

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 56.6585 Seconds
(without alignments)
367.301 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 364
Sequence: 1 PDMDFOSTHPCAFHAAASR.....LNKFTGVIGAFNCGGGMCR 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_minc:*
8: sp_organelle:*
9: sp-phage:*
10: sp_plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	564	100.0	784	10 Q92T62	Q92T62 cucumis sat
2	511	90.6	783	10 Q9FND9	Q9FND9 arabidopsis
3	495	87.8	798	10 Q8VW06	Q8VW06 pisum sativ
4	485	86.0	816	10 Q942T6	Q942T6 oryza sativ
5	409	72.5	857	10 Q9S8X0	Q9S8X0 phaseolus a
6	406	72.0	853	10 Q93XK2	Q93XK2 pisum sativ
7	402	71.3	844	10 Q8RX87	Q8RX87 arabidopsis
8	401	71.1	757	10 Q40077	Q40077 hordeum vul
9	401	71.1	765	10 Q43408	Q43408 brassica o1
10	400	70.9	773	10 Q9SCM1	Q9SCM1 arabidopsis
11	400	70.9	773	10 Q94A08	Q94A08 arabidopsis
12	392	69.5	386	10 Q9M442	Q9M442 cicer ariet
13	388	68.8	863	10 Q8RW08	Q8RW08 stachys aff
14	379	67.2	779	10 Q9M4M7	Q9M4M7 persea amer
15	371	65.8	807	10 Q9STJ4	Q9STJ4 arabidopsis
16	341	60.5	125	10 Q42099	Q42099 arabidopsis

17	213	37.8	1170	10 Q9LFE7	Q9LFE7 arabidopsis
18	183	32.4	649	17 Q97U94	Q97U94 sulfolobus
19	167	29.6	674	17 Q96XG2	Q96XG2 sulfolobus
20	147.5	26.2	702	2 Q93CM6	Q93CM6 bifidobacte
21	115	20.4	357	10 Q93466	Q93466 cicer ariet
22	104.5	18.5	371	10 Q04607	Q04607 arabidopsis
23	80	14.2	283	10 Q9FWM2	Q9FWM2 arabidopsis
24	75	13.3	576	16 Q9ZCM8	Q9ZCM8 rickettsia
25	73	12.9	348	8 Q94W55	Q94W55 coryphopter
26	72	12.8	840	10 Q9LW46	Q9LW46 arabidopsis
27	71	12.6	273	16 Q9PN62	Q9PN62 campylobact
28	71	12.6	382	17 Q970S3	Q970S3 sulfolobus
29	71	12.6	638	5 Q9VYU8	Q9VYU8 drosophila
30	71	12.6	959	12 Q55716	Q55716 chilo iride
31	70.5	12.5	212	2 Q49799	Q49799 mycobacteri
32	70.5	12.5	236	9 Q64059	Q64059 bacterioph
33	70.5	12.5	236	16 Q31962	Q31962 bacillus su
34	70.5	12.5	294	16 Q9S377	Q9S377 mycobacteri
35	69	12.2	615	4 Q9NUX3	Q9NUX3 homo sapien
36	69	12.2	615	4 Q8TAP0	Q8TAP0 homo sapien
37	69	12.2	623	3 Q9P938	Q9P938 plichia stip
38	69	12.2	2023	12 Q91632	Q91632 cherry gree
39	68.5	12.1	183	16 Q911Y7	Q911Y7 pseudomonas
40	68.5	12.1	216	4 Q9BT41	Q9BT41 homo sapien
41	68.5	12.1	219	4 Q9BVW4	Q9BVW4 homo sapien
42	68	12.1	329	17 Q8TW71	Q8TW71 methanopyru
43	68	12.1	2026	12 Q9DVI1	Q9DVI1 cherry gree
44	67	11.9	292	16 Q9AB15	Q9AB15 caulobacter
45	67	11.9	373	5 Q26960	Q26960 trypanosoma

ALIGNMENTS

RESULT 1	Q92T62	PRELIMINARY:	PRT:	784 AA.
AC	Q92T62;			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Raffinose synthase (EC 2.4.1.82).			
GN	RFS.			
OS	Cucumis sativus (Cucumber).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid1; Cucurbitales; Cucurbitaceae; Cucumis.			
OX	NCBI_TaxID=3659;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAF;			
RA	Ohsumi C., Nozaki J., Kida T.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF073744; AAD02832.1; -			
KW	Glycosyltransferase; Transferase.			
SQ	SEQUENCE 784 AA: 86920 MW: 38064491F0908933 CRC64;			
Query Match	100.0%;	Score 564;	DB 10;	Length 784;
Best Local Similarity	100.0%;	Pred. No. 1.7e-54;		
Matches 101;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1 PDMDFOSTHPCAFHAAASRAISGGPIYSDVGKHNFDLKKLVLPQGSIRSEYALP 60			
DB	510 PDMDFOSTHPCAFHAAASRAISGGPIYSDVGKHNFDLKKLVLPQGSIRSEYALP 569			
QY	61 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMCR 101			
DB	570 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMCR 610			
RESULT 2	Q9FND9	PRELIMINARY:	PRT:	783 AA.
ID	Q9FND9			

```
AC Q9FND9;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Raffinose synthase protein.
GN MPO12.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
RX NCBL_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLMBIA;
RC MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones."
RL DNA Res. 4:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006702; BAB11595.1; -
DR EMBL; AY062781; AAL32859.1; -
DR EMBL; AY081645; AAM10207.1; -
SQ SEQUENCE 783 AA; 86237 MW; 3C37DD1D7871888AC CRC64;

Query Match
Best Local Similarity 90.6%; Score 511; DB 10; Length 783;
Matches 90; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 1 PWDMMFQSTHPCAFPAASRAISGPIYSDSYGKHFLLKLVLPDGSILRSEYALP 60
DB 515 PWDMMFQSTHPCAFPAASRAISGPIYSDSYGKHFLLKLVLPDGSILRSEYALP 574
QY 61 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCQGGGMC 101
DB 575 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCQGGGMC 615

RESULT 3
O8YVNG6 PRELIMINARY; PRT; 798 AA.
AC O8YVNG6;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Raffinose synthase (EC 2.4.1.82).
GN RFS.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
RX NCBL_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=CV. WUNDER VON KELVEDON; TISSUE=SEED;
RA Peterbauer T., Mach L., Mucha J., Richter A.;
RT "Molecular characterization of raffinose synthase from pea (Pisum
RT sativum L.) seeds."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A0426475; CAD20127.2; -
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 798 AA; 88717 MW; 8D3F3ED5BF8617B7 CRC64;

Query Match
Best Local Similarity 87.8%; Score 495; DB 10; Length 798;
Matches 85; Conservative 6; Mismatches 9; Indels 0; Gaps 0;

QY 1 PWDMMFQSTHPCAFPAASRAISGPIYSDSYGKHFLLKLVLPDGSILRSEYALP 60
DB 528 PWDMMFQSTHPCAFPAASRAISGPIYSDSYGKHFLLKLVLPDGSILRSEYALP 587
QY 61 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCQGGGMC 100
DB 588 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCQGGGMC 627

RESULT 4
Q942T6 PRELIMINARY; PRT; 816 AA.
AC Q942T6;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Putative raffinose synthase.
GN P0583G08.2.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
RX NCBL_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. NIPPONBARE;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT clone:P0583G08."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP003282; BAB64768.1; -
SQ SEQUENCE 816 AA; 89588 MW; B316EDF3566C5178 CRC64;

Query Match
Best Local Similarity 86.0%; Score 485; DB 10; Length 816;
Matches 83; Conservative 12; Mismatches 6; Indels 0; Gaps 0;

QY 1 PWDMMFQSTHPCAFPAASRAISGPIYSDSYGKHFLLKLVLPDGSILRSEYALP 60
DB 546 PWDMMFQSTHPCAFPAASRAISGPIYSDSYGKHFLLKLVLPDGSILRSEYALP 605
QY 61 TRDCLFEDPLHNGETMLKIMNLKFTGVIGAFNCQGGGMC 101
DB 606 TRDCLFADPLHDGKTMKIMNLKFTGVIGAFNCQGGGMSR 646

RESULT 5
Q9SBZ0 PRELIMINARY; PRT; 857 AA.
AC Q9SBZ0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Galactinol-raffinose galactosyltransferase (EC 2.4.1.67).
OS Phaseolus angularis (Adzuki bean) (Vigna angularis).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
RX NCBL_TaxID=3914;
RN [1]
RP SEQUENCE FROM N.A.
```

DT	01-NOV-1996	(TREMBlrel, 01, Created)
DT	01-NOV-1996	(TREMBlrel, 01, Last sequence update)
DT	01-DEC-2001	(TREMBlrel, 19, Last annotation update)

```

DE Putative imbibition protein.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ALPHA BALOMA; TISSUE=ROOT TIPS;
RA Fujikura Y., Karsen C.K.;
RT "Cauliflower cDNA encoding a putative imbibition protein.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: X79330; CAAS5893.1; -
SQ SEQUENCE 765 AA; 84084 MW; 73398603048E958 CRC64;

Query Match
Best Local Similarity 71.1%; Score 401; DB 10; Length 765;
Matches 67; Conservative 13; Mismatches 21; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYPDGSLRSEYALP 60
DB 476 PDMDMFHSLHPTAEYHAAARAVGCAIYVSDKRGNNHFDLRLKLYLPDGSVLRALPGR 535

QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVGAFCNCGGCMCR 101
DB 536 TRDCLFADPARDGTSLKTIWNNMKFTGIVGVFCNCGAGCMCK 576

RESULT 10
O9SCM1
ID 09SCM1 PRELIMINARY; PRT: 773 AA.
AC 09SCM1;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE Imbibition protein homolog.
GN TBH10.120.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Benes V., Rechmann S., Borkova D., Ansoorge W., Mewes H.W., Lemcke K.,
RA Mayer K.F.X., Queclier F., Salanoubat M.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL133248; CAB6109.1; -
SQ SEQUENCE 773 AA; 85143 MW; 0852F9E67952C8D3 CRC64;

Query Match
Best Local Similarity 70.9%; Score 400; DB 10; Length 773;
Matches 67; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYPDGSLRSEYALP 60
DB 478 PDMDMFHSLHPTAEYHAAARAVGCAIYVSDKRGNNHFDLRLKLYLPDGSVLRALPGR 537

QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVGAFCNCGGCMCR 101
DB 538 TRDCLFADPARDGTSLKTIWNNMKFTGIVGVFCNCGAGCMCK 578

RESULT 11
O94A08
ID 094A08 PRELIMINARY; PRT: 773 AA.
AC 094A08;
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)

```

```

DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Putative imbibition protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,
RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
RA Carinuci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koeseima E., Lam B.,
RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
RA Sakurai T., Satou M., Seki M., Shino P., Southwick A., Tracy S.E.,
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "Full Length cDNA of gene TBH10.120/At3g57520 (GI:5706423).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY050772; AAK92707.1; -
SQ SEQUENCE 773 AA; 85171 MW; 18BDF9E67952C801 CRC64;

Query Match
Best Local Similarity 70.9%; Score 400; DB 10; Length 773;
Matches 67; Conservative 14; Mismatches 20; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYPDGSLRSEYALP 60
DB 478 PDMDMFHSLHPTAEYHAAARAVGCAIYVSDKRGNNHFDLRLKLYLPDGSVLRALPGR 537

QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVGAFCNCGGCMCR 101
DB 538 TRDCLFADPARDGTSLKTIWNNMKFTGIVGVFCNCGAGCMCK 578

RESULT 12
O9M442
ID 09M442 PRELIMINARY; PRT: 386 AA.
AC 09M442;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
DE Putative imbibition protein (Fragment).
OS Cicer arietinum (Chickpea) (Garbanzo).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.
OX NCBI_TaxID=3827;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. CASTELLANA; TISSUE=ETIOLATED OSMOTIC STRESSED EPICOTYLS;
RA Dopolco B., Romo S., Labrador E.;
RT "A putative imbibition protein is expressed in chickpea epicotyls.";
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ271668; CAB71135.1; -
FT NON_TER
SQ SEQUENCE 386 AA; 42180 MW; 357AF9FBC8D71650 CRC64;

Query Match
Best Local Similarity 69.5%; Score 392; DB 10; Length 386;
Matches 68; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDSVGKHNFDLKLKLYPDGSLRSEYALP 60
DB 90 PDMDMFHSLHPTAEYHAAARAVGCAIYVSDKRGNNHFDLKLKLYLPDGSVLRALPGR 149

QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVGAFCNCGGCMCR 101
DB 150 TRDCLFVDPARDGTSLKTIWNNMKFTGIVGVFCNCGAGCMCK 190

RESULT 13
O8RW08

```



```

ID 08RW08 PRELIMINARY: PRT: 863 AA.
AC 08RW08;
DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Stachyose synthase (EC 2.4.1.67).
GN STS.
OS Stachys affinis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; eusterids I; Lamiales; Lamiales; Stachys.
OX NCBI_TaxID=168825;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Pesch M., Schmitz K.;
RT "Molecular cloning of a cDNA encoding for stachyose synthase from
RT Stachys alba0111";
RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ344091; CAC86963.1; -
KM Transferase: Glycosyltransferase.
SQ SEQUENCE 863 AA; 95227 MW; DD96FE66099DEFD CRC64;

Query Match 68.8%; Score 388; DB 10; Length 863;
Best Local Similarity 68.7%; Pred. No. 9.9e-35;
Matches 68; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

OY 1 PDWDMFQSTHPCAFHNASRAISGPIYVSDSGKHNFDLKKLVPDGSILRSEYALP 60
DB 593 PDWDMFQSDHCHSKAFHNASRAISGPIYVSDSGKHNFDLKKLVPDGSILRSEYALP 652
OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCOGGAGW 99
DB 653 TRDCLFKNPLFDSKTLIKINLNKFTGVIGAFNCOGGAGW 691

RESULT 14
OY 09M4M7 PRELIMINARY: PRT: 779 AA.
AC 09M4M7;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 19, Last annotation update)
DE Putative seed imbibition protein.
GN SIP.
OS Persa americana (Avocado).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. HASS; TISSUE=MESOCARP;
RA Zamorano J.P., Evans A.D., Dopico B., Lowe A.L., Wilson I.D.,
RA Merodio C., Giletson D.;
RT "Isolation and characterization of cDNAs for mRNAs regulated during
RT cold storage of avocado (Persea americana Mill.) fruit.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ133148; CAB77245.1; -
SQ SEQUENCE 779 AA; 85368 MW; C3A8B43160316785 CRC64;

Query Match 67.2%; Score 379; DB 10; Length 779;
Best Local Similarity 64.4%; Pred. No. 8.9e-34;
Matches 65; Conservative 12; Mismatches 24; Indels 0; Gaps 0;

OY 1 PDWDMFQSTHPCAFHNASRAISGPIYVSDSGKHNFDLKKLVPDGSILRSEYALP 60
DB 482 PDWDMFQSHLPAEYHGAAYAGVCGPIYVSDSGKHNFDLKKLVPDGSILRSEYALP 541
OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCOGGAGW 101
DB 542 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCOGGAGW 582

```

```

RESULT 15
OY 09SYJ4 PRELIMINARY: PRT: 807 AA.
AC 09SYJ4;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
DE Putative raffinose synthase OR seed IMBIBITION protein.
GN T7B11.23 OR AT4G01970.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Huang E.N., Nascimeto L., de la Bastide M., Habermann K., Vil M.D.,
RA Preston R.R., Spielgel L.A., See L.H., Shah R., Matero A.,
RA O'Shaughnessy A., Rodriguez M., Shekher M., Swaby I., Schutz K.,
RA Parnell L.D., Dedhia N.N., McComble W.R.;
RT "Arabidopsis thaliana BAC T7B11 from chromosome IV near 10 cm.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Spielgel L.A., Huang E.N., Nascimeto L.U., de la Bastide M., Vil D.M.,
RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007138; AND22659.1; -
DR EMBL; AL161493; CAB80690.1; -
SQ SEQUENCE 807 AA; 90122 MW; 963DCD5A827B338B CRC64;

Query Match 65.8%; Score 371; DB 10; Length 807;
Best Local Similarity 67.3%; Pred. No. 7.3e-33;
Matches 68; Conservative 13; Mismatches 18; Indels 2; Gaps 1;

OY 1 PDWDMFQSTHPCAFHNASRAISGPIYVSDSGK--HNFDLKKLVPDGSILRSEYALP 58
DB 530 PDWDMFQSDHCHSKAFHNASRAISGPIYVSDSGKASHNFDLKKLAFPGSTIPRCVHYA 589
OY 59 LPTRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCOGGAGW 99
DB 590 LPTRDCLFKNPLFDSKTLIKINLNKFTGVIGAFNCOGGAGW 630

```

Search completed: March 27, 2003, 10:06:04
 Job time : 59.6585 secs

Dd 476 PDMDFSHLHTAETAEYHAAARVAGGCAIIVSDKPCNNHFDDLRLKVLVPDGSYLRRALPGRP 535

Qy 61 TRDCLEFDPLINGETMLKIYNLNKFYIGAFNQGGGWCR 101
||||| : : : ||||| : : ||||| : :
Db 536 TRDCLFADPARPDGTSILKIYNMKNFTGLGVGFNQGCGMWCK 576

RESULT 3
T46188
imidation protein homolog - Arabidopsis thaliana
N:Alternate names: protein T8H10.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46188
R:Beenes, V.; Rechmann, S.; Borkova, D.; Ansoyge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.F. H.
submitted to the Protein Sequence Database, January 2000
.:Reference number: 223014

A:Cross-references: EMBL:AL33248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:Introns: 64/2; 146/1; 176/2; 192/3; 223/2; 259/2; 300/2; 484/3; 507/2; 552/2; 625/1; 625/2; 625/3; 625/4; 625/5; 625/6; 625/7; 625/8; 625/9; 625/10; 625/11; 625/12; 625/13; 625/14; 625/15; 625/16; 625/17; 625/18; 625/19; 625/20; 625/21; 625/22; 625/23; 625/24; 625/25; 625/26; 625/27; 625/28; 625/29; 625/30; 625/31; 625/32; 625/33; 625/34; 625/35; 625/36; 625/37; 625/38; 625/39; 625/40; 625/41; 625/42; 625/43; 625/44; 625/45; 625/46; 625/47; 625/48; 625/49; 625/50; 625/51; 625/52; 625/53; 625/54; 625/55; 625/56; 625/57; 625/58; 625/59; 625/60; 625/61; 625/62; 625/63; 625/64; 625/65; 625/66; 625/67; 625/68; 625/69; 625/70; 625/71; 625/72; 625/73; 625/74; 625/75; 625/76; 625/77; 625/78; 625/79; 625/80; 625/81; 625/82; 625/83; 625/84; 625/85; 625/86; 625/87; 625/88; 625/89; 625/90; 625/91; 625/92; 625/93; 625/94; 625/95; 625/96; 625/97; 625/98; 625/99; 625/100; 625/101; 625/102; 625/103; 625/104; 625/105; 625/106; 625/107; 625/108; 625/109; 625/110; 625/111; 625/112; 625/113; 625/114; 625/115; 625/116; 625/117; 625/118; 625/119; 625/120; 625/121; 625/122; 625/123; 625/124; 625/125; 625/126; 625/127; 625/128; 625/129; 625/130; 625/131; 625/132; 625/133; 625/134; 625/135; 625/136; 625/137; 625/138; 625/139; 625/140; 625/141; 625/142; 625/143; 625/144; 625/145; 625/146; 625/147; 625/148; 625/149; 625/150; 625/151; 625/152; 625/153; 625/154; 625/155; 625/156; 625/157; 625/158; 625/159; 625/160; 625/161; 625/162; 625/163; 625/164; 625/165; 625/166; 625/167; 625/168; 625/169; 625/170; 625/171; 625/172; 625/173; 625/174; 625/175; 625/176; 625/177; 625/178; 625/179; 625/180; 625/181; 625/182; 625/183; 625/184; 625/185; 625/186; 625/187; 625/188; 625/189; 625/190; 625/191; 625/192; 625/193; 625/194; 625/195; 625/196; 625/197; 625/198; 625/199; 625/200; 625/201; 625/202; 625/203; 625/204; 625/205; 625/206; 625/207; 625/208; 625/209; 625/210; 625/211; 625/212; 625/213; 625/214; 625/215; 625/216; 625/217; 625/218; 625/219; 625/220; 625/221; 625/222; 625/223; 625/224; 625/225; 625/226; 625/227; 625/228; 625/229; 625/230; 625/231; 625/232; 625/233; 625/234; 625/235; 625/236; 625/237; 625/238; 625/239; 625/240; 625/241; 625/242; 625/243; 625/244; 625/245; 625/246; 625/247; 625/248; 625/249; 625/250; 625/251; 625/252; 625/253; 625/254; 625/255; 625/256; 625/257; 625/258; 625/259; 625/260; 625/261; 625/262; 625/263; 625/264; 625/265; 625/266; 625/267; 625/268; 625/269; 625/270; 625/271; 625/272; 625/273; 625/274; 625/275; 625/276; 625/277; 625/278; 625/279; 625/280; 625/281; 625/282; 625/283; 625/284; 625/285; 625/286; 625/287; 625/288; 625/289; 625/290; 625/291; 625/292; 625/293; 625/294; 625/295; 625/296; 625/297; 625/298; 625/299; 625/300; 625/301; 625/302; 625/303; 625/304; 625/305; 625/306; 625/307; 625/308; 625/309; 625/310; 625/311; 625/312; 625/313; 625/314; 625/315; 625/316; 625/317; 625/318; 625/319; 625/320; 625/321; 625/322; 625/323; 625/324; 625/325; 625/326; 625/327; 625/328; 625/329; 625/330; 625/331; 625/332; 625/333; 625/334; 625/335; 625/336; 625/337; 625/338; 625/339; 625/340; 625/341; 625/342; 625/343; 625/344; 625/345; 625/346; 625/347; 625/348; 625/349; 625/350; 625/351; 625/352; 625/353; 625/354; 625/355; 625/356; 625/357; 625/358; 625/359; 625/360; 625/361; 625/362; 625/363; 625/364; 625/365; 625/366; 625/367; 625/368; 625/369; 625/370; 625/371; 625/372; 625/373; 625/374; 625/375; 625/376; 625/377; 625/378; 625/379; 625/380; 625/381; 625/382; 625/383; 625/384; 625/385; 625/386; 625/387; 625/388; 625/389; 625/390; 625/391; 625/392; 625/393; 625/394; 625/395; 625/396; 625/397; 625/398; 625/399; 625/400; 625/401; 625/402; 625/403; 625/404; 625/405; 625/406; 625/407; 625/408; 625/409; 625/410; 625/411; 625/412; 625/413; 625/414; 625/415; 625/416; 625/417; 625/418; 625/419; 625/420; 625/421; 625/422; 625/423; 625/424; 625/425; 625/426; 625/427; 625/428; 625/429; 625/430; 625/431; 625/432; 625/433; 625/434; 625/435; 625/436; 625/437; 625/438; 625/439; 625/440; 625/441; 625/442; 625/443; 625/444; 625/445; 625/446; 625/447; 625/448; 625/449; 625/450; 625/451; 625/452; 625/45

Query Match	70.98;	Score 400;	DB 2;	Length 773;
Best Local Similarity	66.38;	Pred. No. 1e-34;		
Matches 67;	Conservative 14;	Mismatches 20;	Indels 0;	Gaps 0;

```

QY      1 PDWMDFQSTHPCAAFHAAASRAISGGPIYVSDSVGKHNFDLLKLVLPDGSILRSEYYALP 60
          |||||  |  |  :||:|:|  |  |||||  |  |||||:|:|:|:|:|:|:|:|:|  |
Db     478 PDWMDFHSLHPTAEYHAAARAVGCAIYVSDKPGNNHFDLLKLVLPDGSVLRKALPGRP 53

```

```

0y      61  TRDCLFEDPLHNGETMLKINWLNKFTGVIGAFNCOGGGKCR 101
          ||||| | : : |||||:||||:| |||| ||:
Db      538  TRDCLFADPRADGISLTKIMNMNFTGIVGVFNCGAGCKCK 578

```

RESULT 4
 C85025
 hypothetical protein AT4g01970 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001
 C:Accession: C85025
 R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory, New York, NY, USA
 Nature 402, 769-777, 1999
 A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
 #:Reference number: A85001; MUID:20083488; PMID:10617198

A:Cross-references: GB:NC_001268; NID:g7266581; PIDB:CAB80690.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g01970
A:Map position: 4

Query Match	65.8%;	Score 371;	DB 2;	Length 807;
Best Local Similarity	67.3%;	Pred. NO. 1.4e-31;		
Matches 68;	Conservative 13;	Mismatches 18;	Indels 2;	Gaps 1;

Qy 1 PDWMFQSTHPCAAFFHAASRAISGPGPIYVSDYCK--HNFDLKKLVLPDGSILRSEYYA 58
||||||| | : ||||| | : : || ||||| : ||
Db 530 PDWMFQSDHYCAEYHAASRAISGPGPIYVSDHLGKASHNFDLKKLAFFGSTIPRCVHYA 589

QY 59 LPTRDCLEFEDPLHNGETMLKIMNLKFTVGIGAFNCOGGW 99
|||||:::|:|:::||||| ||| ||||| |||
DB 590 LPTRDSLEFKNPLDEDSILKIFENFKFGVIGITFNCOGAGW 630

RESULT 5
C96599

C:Species: *Arabidopsis thaliana* (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001
C:Accession: C96539
R:Thellogis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alor-Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,ansen, N.F.; Hughes, B.; Huizaf, L.
Nature 408, 815-820, 2000

A: Authors: Huerf J.L.; Jenkins, J.S.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Llo, Z.A.; Lueros, J.S.; Malli, R.; Martzfeldt, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A: Authors: Salzedo, S.L., Schwartz, J.R., Shinn, P., Southwick, A.M., Sun, H., Tallal, K.E., Wu, D., Yu, G., Fraser, C.M., Venter, J.C., Davis, R.W.
Article: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.
A:Reference number:A68141; MUID:21016719; PMID:11130712

A:Cross-references: GB:AE005173; NID:98778496; PIDN:AAF79504.1; GSPDB:GN001411
C:Genetics:
A:Gene: F20N2.14
A:Map position: 1

Query Match 37.8%; Score 213; DB 2; Length 1170;

Matches 39; Conservative 12; Mismatches 15; Indels 0; Gaps 0;

Db	498	PDMDHSLHMAEYHAARAVGGCAIVSDPKGQHDNFNLRLKLVLRDGSILRAKLPRGR	557
QY	61	TRDCLF	66

Db 558 TRELVEY 563

RESULT 6
 D90496
 hypothetical protein SSO3127 [imported] - *Sulfolobus solfataricus*
 C:Species: *Sulfolobus solfataricus*
 C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 C:Accession: D90496
 R:She, O.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Ch
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redde
 arreft, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
 submitted to GenBank, April 2001
 A:Description: *Sulfolobus solfataricus* complete genome.
 A:Reference number: A99139

A:Cross-References: GB:AE006641; MID:g13816550; PIDN:AAK43227.1; GSPDB:GN001555
C:Genetics:
A:Gene: SSO3127

Query Match	32.4%	Score 183;	DB 2;	Length 649;
Best Local Similarity	41.6%;	Pred. No. 1,5e-11;		
Matches 42;	Conservative 14;	Mismatches 37;	Indels 8;	Gaps 3;

```

QY      1 PDMDWEGSTHCCAFPHASRAISGGPIYVSISYGHK---NEDDLKLTLYPDGSLIRSEY 56
      ||:||||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      455 PDYDMFMSYDPYAKVHLVARVSSGPIYITD---RHPERTNIELLRMAVLNPGSVIRVDE 511

```

QY 57 YALPTROCLFEDPLHNGETMLKIYNLKFETGVIAGAFNQQGG 97
||| ||| : || |
Db 512 PALITDELLFPDLRE-RVLLKIKCKVKGNATFFNLNSG 551

RESULT 7

[illegible]

```

C:Genetics:
A:Gene: aq_1075

Query Match
Best Local Similarity 13.6%; Score 76.5; DB 2; Length 165;
Matches 28; Conservative 12; Mismatches 36; Indels 19; Gaps 6;

Oy 1 PDMFMFQSTHCCAFHAASRAISGPIYVSDS--VGKHNFDLKKLV---LPDGSIL--R 53
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 82 PKWIMFAARLPLNGKRLHK-----SGGKVRDAVIVYGEQVNELENPFISFKYSDGRITNFR 137

Oy 54 SEYYALPTROCLFEDPLHNGETMLKIMNLNKTGCV 88
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 IEEF----RYLRKSPKYGPSFRSEYW---FICV 164

RESULT 10
B71676
Probable ABC transporter mbsbA2 - Rickettsia prowazekii
C:Species: Rickettsia prowazekii
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 03-Nov-2000
C:Accession: B71676
R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sichteritz-Ponten, T.; Alismark
Nature 396, 133-140, 1998
A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
A:Reference number: A71630; MUID:99039499; PMID:9823893
A:Accession: B71676
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-576 <AND>
A:Cross-references: GB:AJ235272; GB:AJ235269; NID:93861033; PIDN:CA15132.1; PID:9386
A:Experimental source: strain Madrid E
C:Genetics:
A:Gene: mbsbA2; RP696
C:Superfamily: Rickettsia prowazekii probable ABC transporter mbsbA2; ATP-binding class
C:Keywords: ATP
F:354/Domain: ATP-binding cassette homology <ABC>

Query Match
Best Local Similarity 13.3%; Score 75; DB 1; Length 576;
Matches 27; Conservative 14; Mismatches 25; Indels 34; Gaps 5;

Oy 15 FHASRAISGPIYVSDSVGKHNFDLKKLVLP-----DGSILRSEYALPTR 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 282 YVAILAGVSSGCI-----FELISEHMLPTTALEIRITIIDKTSIVHNNYVALNNS 331

Oy 63 DCL--FE--DPLNGETMLKI-----WNLNFTGIG 90
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 332 DAISTEFKNVDTYNSRPNLKVINMSLKNKFGIVG 371

RESULT 11
E81330
Probable periplasmic protein Cj1235 [imported] - Campylobacter jejuni (strain NCTC 11)
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C:Accession: E81330
R:Packhill, J.; Wren, B.W.; Mungall, K.; Kealey, J.M.; Churcher, C.; Basham, D.; Chilling
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barr
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A:Reference number: A81250; MUID:20150912; PMID:10688204
A:Accession: E81330
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-273 <PAR>
A:Cross-references: GB:AL139077; GB:AL111168; NID:96968444; PIDN:CAB73489.1; PID:9696
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: Cj1235

Query Match
Best Local Similarity 12.6%; Score 71; DB 2; Length 273;
Matches 26; Conservative 12; Mismatches 36; Indels 19; Gaps 6;

```

Matches 25; Conservative 17; Mismatches 40; Indels 14; Gaps 4;

QY 13 AAFHAAS--RAISGPIYSDS---VGKHNFDLKKLVLPDGSILRSEYVALPTRCLF 66

Db 165 ASHSGTDRATGTPGTPIYANSGVAKIANDRYAGNSGVYDHFGIYSQYHLKIDVKI 224

QY 67 EDPLHNGETMLKIMNKTGIGA---FNCGGG 98

Db 225 GQIKKGGELI---GLSGATGVRVSCPHLHFGILAG 256

RESULT 12

T03053

ribonucleoside-diphosphate reductase large chain homolog - Chilo iridescent virus

C:Species: Chilo iridescent virus

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 08-Oct-1999

C:Accession: T03053

R:Bahr, U.; Tidona, C.A.; Darai, G.

A:Title: The DNA sequence of Chilo iridescent virus between the genome coordinates 0.101

A:Reference number: Z14834; MUID:98141693; PMID:9482589

A:Accession: T03053

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-959 <BAH>

A:Cross-references: EMBL:AF003534; NID:92738385; PIDN:AAB94427.1; PID:92738400

Query Match 12.6%; Score 71; DB 2; Length 959;

Best Local Similarity 13.3%; Pred. No. 24; Mismatches 29; Indels 8; Gaps 4;

Matches 25; Conservative 13; Mismatches 29; Indels 8; Gaps 4;

QY 19 SRAISGP-IYSDSVGKHNFDLKKLVLPDGSILRSEYVALPTRCLFED---PLHNGE 74

Db 236 TRMKTGNPTICFDTCNKHMYDFQKKGL---TIKSNLCVADETMLTFDGGFPIKIDLE 292

QY 75 -TMLKIMNKTGTV 88

Db 293 GKTIKWNENESSV 307

RESULT 13

S72873

hypothetical protein B2126_F2-70 - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001

C:Accession: S72873

R:Smith, D.R.; Robison, K.

A:Description: Mycobacterium leprae cosmid B2126.

A:Reference number: S72585

A:Accession: S72873

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-212 <SMT>

A:Cross-references: EMBL:U00017; NID:9466994; PIDN:AAAI7213.1; PID:9467028

C:Genetics:

A:start codon: GTG

Query Match 12.5%; Score 70.5; DB 2; Length 212;

Best Local Similarity 34.8%; Pred. No. 4.7; Mismatches 18; Indels 17; Gaps 5;

Matches 23; Conservative 8; Mismatches 18; Indels 17; Gaps 5;

QY 41 LKKLVLPDGSILRSEYVALPTRCLFEDPLHNGETMLKIMNKTGIVGAFN-----C 94

Db 122 LRLIYVADSQLD---YA-PDRDEL---LREGKTLMAIWRITQSGEGDPRFNRSLC 173

QY 95 QGGGKC 100

Db 174 ---GWC 176

RESULT 14

T12810

hypothetical protein yomv - Bacillus subtilis phage SPBc2

C:Species: Bacillus subtilis phage SPBc2

C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 15-Oct-1999

C:Accession: T12810; F69912

R:Lazarevic, V.; Duesterhoeft, A.; Soldo, B.; Hilbert, H.; Mausel, C.; Karamata, D.

A:Description: The complete nucleotide sequence of the Bacillus subtilis SPbetac2 pro

A:Reference number: Z17583

A:Accession: T12810

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-236 <LAZ>

A:Cross-references: EMBL:AF020713; NID:93025478; PID:93025524; PIDN:AAIC13019.1

A:Authors: Foulger, D.; Fultz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lechi, J.; Harwood, C.R.; Henault, A.; Hilbert, H.; Holstappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadate, Y.; Sato, T.; Scagl

A:Authors: Schlach, S.; Schroeter, R.; Scorfone, F.; Sekiguchi, J.; Sekowska, A.; Se

akuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtil

A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: F69912

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-236 <KUN>

A:Cross-references: GB:Z99115; GB:AL009126; NID:92634478; PIDN:CAB14039.1; PID:el1835

A:Experimental source: strain 168

C:Genetics:

A:gene: yomv

Query Match 12.5%; Score 70.5; DB 2; Length 236;

Best Local Similarity 21.5%; Pred. No. 5.3; Mismatches 18; Mismatches 35; Indels 31; Gaps 4;

Matches 23; Conservative 18; Mismatches 35; Indels 31; Gaps 4;

QY 4 DMFQSTHPCAA-FHAASRAISGPIYSDS---VGKHNFDLKKLVLPDGSILRSEYVAL 59

Db 70 DLVEIHSVLPDRHEAVFYKFSKVLINKDSLILGKHNFPRLKIL- 116

QY 60 PTRDCLFEDPLHNGETMLKIMNKTGIVGAFN-----FTGVGAFN 93

Db 117 -DMHCITEDKVDNDELQEPHDSLSLKQODSOSDLKDIYCVAAFN 162

RESULT 15

B87073

conserved hypothetical protein M1312 [imported] - Mycobacterium leprae

C:Species: Mycobacterium leprae

C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001

C:Accession: B87073

R:Coile, S.T.; Eigmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;

A:Authors: Rutler, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;

A:Title: Massive gene decay in the leprosy bacillus.

A:Reference number: A86909; MUID:21128732; PMID:11234002

A:Accession: B87073

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-294 <STO>

A:Cross-references: GB:AL450380; NID:g13093231; PIDN:CAC31693.1; GSPDB:GN00147

C:Genetics:

A:Gene: M1312

Query Match 12.5%; Score 70.5; DB 2; Length 294;

T12810

Fri Apr 4 12:31:35 2003

us-09-675-208-5_copy_510_610.rpr

Page 5

Best Local Similarity 34.8%; Pred. NO. 6.9;
Matches 23; Conservative 8; Mismatches 18; Indels 17; Gaps 5;

41 LKKLVLPDGSILRSEYYALPTRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFN-----C 94

Db 204 LRLIYLADSQLD---YA-PDRDEL----LRFCKTLMAINRIQFSGETGDFRPNPSRLC 255

QY 95 QGGGWC 100
111

Db 256 ---GWC 258

Search completed: March 27, 2003, 10:06:54
Job time : 30.9451 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 16.628 seconds

(without alignments)
251.930 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610
Perfect score: 564
Sequence: 1 PDMDFGQSTHPCAFHAAASR.....LNKFTGVIGAFNCQGGWCR 101

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	76.5	13.6	165	YA75_AQUAE	067169 aquifex aeo
2	69.5	12.3	544	SYE_AERPE	0949h1 aeropyrum p
3	67	11.9	258	FLA3_PYRKO	09v2w9 pyrococcus
4	66.5	11.8	149	MRA2_TREPA	083398 treponema p
5	64.5	11.4	358	YJ9L_YEAST	P47175 saccharomyc
6	64.5	11.4	626	PM17_MOUSE	060696 mus musculu
7	64.5	11.4	667	TS11_GIALA	003385 giardia lam
8	64	11.3	105	GLPE_HAEIN	P44819 haemophilus
9	63.5	11.3	402	VN34_ROTPO	P27586 porcine rot
10	63.5	11.3	721	BBS2_RAT	0195m9 ratulus norv
11	63.5	11.3	819	Y083_NPVOP	010336 orgyia pseu
12	63	11.2	884	TERT_YEAST	006163 saccharomyc
13	63	11.2	982	NIA_NEUCR	P08619 neurospora
14	62.5	11.1	165	Y872_RICPR	092694 rickettsia
15	62.5	11.1	220	AP12_SOUTU	043646 solanum tub
16	62.5	11.1	263	YC56_PORPU	P51308 porphyra pu
17	62.5	11.1	261	SUHB_BUCAI	P57372 buchiera ap
18	62.5	11.1	481	HTR3_HALNI	09h000 halobacteri
19	62.5	11.1	481	HTR3_HALSA	048316 halobacteri
20	62	11.0	617	VAAT_HUMAN	P38606 homo sapien
21	62	11.0	993	ROXN_HUMAN	Q949r2 homo sapien
22	62	11.0	1264	YGN3_YEAST	P53125 saccharomyc
23	61.5	10.9	295	DAP1_BACHD	09K632 bacillus ha
24	61.5	10.9	346	CCPR_PSEAE	P14532 pseudomonas
25	61	10.8	590	VPP_BPP2	P25479 bacterioph
26	60.5	10.7	249	RNPI_YEAST	P32385 saccharomyc
27	60.5	10.7	293	MOTY_VIBPA	P46233 vibrio para
28	60.5	10.7	340	CELF_VZVD	P09261 varicella-z
29	60.5	10.7	715	PAP1_XENLA	P51004 xenopus lae
30	60.5	10.7	1044	IFZP_PYRHO	058822 pyrococcus
31	60	10.6	210	RHOD_HUMAN	000212 homo sapien
32	60	10.6	235	SOML_GADMO	P21919 gadus morhu
33	60	10.6	294	DAPA_RICCN	Q92125 rickettsia

34	60	10.6	294	DAPA_RICMO	Q9akg3 rickettsia
35	60	10.6	294	DAPA_RICRI	Q9akj9 rickettsia
36	60	10.6	765	TRPG_PHYBL	P20409 p anthranil
37	59.5	10.5	188	AP18_SOLTU	P16348 solanum tub
38	59.5	10.5	220	AP18_SOLTU	P17979 solanum tub
39	59.5	10.5	358	T2H2_HAEPA	P36433 haemophilus
40	59.5	10.5	423	MTN4_NEDLA	P50182 neisseria 1
41	59.5	10.5	525	Y487_TREPA	083500 treponema p
42	59.5	10.5	529	Y119_NPVOP	010358 orgyia pseu
43	59.5	10.5	717	YM4_CABEL	P34520 caenorhabdl
44	59	10.5	377	AMPC_ECOLI	P00811 escherichia
45	59	10.5	390	ARP2_SCHPO	Q9uu11 schizosacch

ALIGNMENTS

RESULT 1	YA75_AQUAE	STANDARD;	PRT;	165 AA.
ID	YA75_AQUAE			
AC	067169;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Hypothetical protein AO_1075.			
GN	AO_1075.			
OS	Aquifex aeolicus.			
OC	Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;			
OX	Aquifex.			
NCBI_TaxID=63363;				
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=VF5;			
RX	MEDLINE=98196666; PubMed=9537320;			
RA	Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,			
RA	Graham D.E., Overbeek R., Sneed M.A., Aujeay M., Huber R.,			
RA	Feldman R.A., Short J.M., Olson G.J., Swanson R.V.?			
RT	"The complete genome of the hyperthermophilic bacterium Aquifex			
RT	aeolicus.";			
RL	Nature 392:353-358(1998).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement. (See http://www.isb-sdb.ch/announce/			
CC	or send an email to license@sdb.ch).			
CC	-----			
DR	EMBL: AEO00721; AAC07132.1; -			
KW	Hypothetical protein; Complete proteome.			
SQ	SEQUENCE 165 AA; 19532 MW; 027AB8C916795B18 CRC64;			
Query Match	13.6%; Score 76.5; DB 1; Length 165;			
Best local Similarity	29.5%; Pred. No. 0.22;			
Matches	28; Conservative 12; Mismatches 36; Indels 19; Gaps 6;			
QY	1 PDMDFGQSTHPCAFHAAASRAISGPIYSDS--VGKHNFDLLKIV---LPDGSIL--R 53			
DB	82 PKWIMFAARLPNKLEHPK---SGGKYREDVAIVGEGNLEPNFTSEKYSQGRILNFR 137			
QY	54 SEYVALPTRDCLFEDPLNGETMLKIWNLNKFTGV 88			
DB	138 IERY----RYLRKSKRKYGSPFSFEW---FIGV 164			
RESULT 2	SYE_AERPE	STANDARD;	PRT;	544 AA.
ID	SYE_AERPE			
AC	0949h1;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			

DE glutamyl-tRNA synthetase (EC 6.1.1.17) (Glutamate--tRNA ligase)
DE (GURS).
GN GLTX OR APE2317.
OS Aeropyrum pernix.
OC Archaea: Crenarchaeota: Thermoprotei: Desulfurococcaceae:
OC Desulfurococcaceae: Aeropyrum.
ON NCBI_TaxID=56636;
OX [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1.
RX MEDLINE=99310339; PubMed=10382966;
RA Kawabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anka H., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kishida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1";
RL DNA Res. 6:83-101(1999).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + tRNA(Glu) = AMP +
CC diphosphate + L-glutamyl-tRNA(Glu).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO CLASS-I AMINOACYL-TRNA SYNTHETASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000064; BAA81329.1; -
DR HSSP: P00962; 1GTR
DR InterPro: IPR004526; GLX_arch.
DR InterPro: IPR000924; Glu_tRNA-synt_1c.
DR InterPro: IPR001412; tRNA-synt_1.
DR Pfam: PF00749; tRNA-synt_1c; 1.
DR PRINTS: PR00987; TRNASYNTHGLD.
DR TIGRFAMs: TIGR00463; glx_arch; 1.
DR PROSITE: PS00178; AA_TRNA_LIGASE_I; FALSE_NEG.
KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
KW Complete proteome.
FT SITE 79 89 "HIGH" REGION.
FT SQT 544 AA; 62894 MW; 52B85CDB3EEA84 CRC64;
SQ SEQUENCE 544 AA; 62894 MW; 52B85CDB3EEA84 CRC64;
Query Match 12.3%: Score 69.5; DB 1; Length 544;
Best Local Similarity 35.9%: Pred. No. 5;
Matches 28; Conservative 5; Mismatches 30; Indels 15; Gaps 4;
QY 3 WD--MFQSTHPCAFAHSAISGPIYVSDSGKHNFDLLKLVDPDGSILRSEYVALP 60
DB 140 WDEYIOSLMEVFYSVARAIERGCAVY-DNCRGREGKELLS-----KGEY--CP 186
QY 61 TRDCLFEDPLHNGETMLK 78
DB 187 TRDLPEDNLEFEKMLE 204
RESULT 3
FL3_PYRKO STANDARD; PRT; 258 AA.
AC 09V2W9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellin B3 precursor.
GN FLA33.
OS Pyrococcus kodakaraensis.
OC Archaea: Euryarchaeota: Thermococci: Thermococcales; Thermococcaceae;
OC Thermococcus.
OX NCBI_TaxID=69014;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOD1.
RX MEDLINE=99413236; PubMed=10483738;
RA Nagahisa K., Ezaki S., Fujiwara S., Imanaka T., Takagi M.;
RT "Sequence and transcriptional studies of five clustered flagellin
RT genes from hyperthermophilic archaean Pyrococcus kodakaraensis KOD1";
RL FEMS Microbiol. Lett. 178:183-190(1999).
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF FLAGELLA.
CC -1- SIMILARITY: BELONGS TO THE ARCHAEAL FLAGELLIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB018434; BAA84107.1; -
DR InterPro: IPR002774; Arch_flagellin.
DR Pfam: PF01917; Arch_flagellin; 1.
KW Flagella; Multigene family.
FT PROPEP 1 8 BY SIMILARITY.
FT CHAIN 9 258 FLAGELLIN B3.
SQ SEQUENCE 258 AA; 27443 MW; E0AD08C1B7ABBEAC CRC64;
Query Match 11.9%: Score 67; DB 1; Length 258;
Best Local Similarity 28.3%: Pred. No. 4.1;
Matches 26; Conservative 15; Mismatches 29; Indels 22; Gaps 4;
QY 20 RAISGPIYVSDSGKHNFDLLK-LVLPDGSI-----TRSEYVALPTR 62
DB 77 KPISKLAIVSPNAGSGIDMKKRYVLSDGSTEAVLKYTONSDSDGTLDKQVFAVGMP 136
QY 63 DCLFED---PLHNGETMLKIMLN-KFTGVI 89
DB 137 DNVFEDDTGTTAADGDQYITWSELNDKTFGLI 168
RESULT 4
MRZ_TREPA STANDARD; PRT; 149 AA.
AC 083398;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Protein mrz.
GN MRZ OR TP0383.
OS Treponema pallidum.
OC Bacteria: Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nichols;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod W.P., Salzberg S., Patereson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith R.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete";
RL Science 281:375-388(1998).
CC -1- SIMILARITY: BELONGS TO THE MRZ FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

FT VARIANT 175 175 R -> G (IN SILVER).
 FT VARIANT 373 373 D -> N (IN SILVER).
 FT VARIANT 471 471 F -> S (IN SILVER).
 FT VARIANT 603 626 AAPASGLRARGIENSPILGQOV -> SSASLRSSRPBW
 FT SEQUENCE 626 AA: 65980 MW: 7AB941D2E3FB1044 CRC64;
 SO

Query Match 11.4%; Score 64.5; DB 1; Length 626;
 Best Local Similarity 24.8%; Pred. No. 21;
 Matches 29; Conservative 17; Mismatches 42; Indels 29; Gaps 7;

QY 1 PDDMROSTHPCAFHAAASRAISGPIYV-----SDSYGKINFDLKKLVLPDGSIL----- 52
 DB 50 PEWTEVQGSNCWRGGVSLVINDGPTLVGAMASFSIALH-FPGSQK-VLPDQGVIMANN 107
 OY 53 -----RSEYVALPTRDCLFED--PLHNG-----ETMLKIMVL-WKFTGVIG 90
 DB 108 TIINGSGVWGQVRYOEPDADCVFPDGGCPGSPGPKRSPFYVWTKWGIWQVLG 164

RESULT 7

TS1L_GIALA STANDARD; PRT: 667 AA.
 ID TS1L_GIALA 003185;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Major surface trophozoite antigen 11 precursor.
 GN TSP11.
 OS Giardia lamblia (Giardia intestinalis).
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardinae; Giardia.
 OX NCBI_TaxID=5741;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate AD-1;
 RX MEDLINE=93241215; PubMed=8479449;
 RA Ey P.L., Khanna K., Manning P.A., Mayrhofer G.
 RT "A gene encoding a 69-kilodalton major surface protein of Giardia
 intestinalis trophozoites";
 RL Mol. Biochem. Parasitol. 58:247-258(1993).
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ANCHORED TO THE
 CC PLASMA MEMBRANE.
 CC -1- DOMAIN: CONTAINS 27 REPEATS OF A CXXC MOTIF.
 CC -1- SIMILARITY: BELONGS TO THE GIARDIA VARIANT SURFACE PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: M95814; AAA02687.1; -
 CC PIR: A48579; A48579.
 DR HSSP: P02468; IKLO.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR005127; Giardia_VSP.
 DR Pfam: PF03302; VSP; 1.
 DR SMART: SM00181; EGF; 3.
 DR SMART: SM00261; FUF; 5.
 DR Antigen; Repeat; Transmembrane; Signal.
 KW SIGNAL
 FT CHAIN 18 667 MAJOR SURFACE TROPHOZOITE ANTIGEN 11.
 FT DOMAIN 18 663 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 634 662 POTENTIAL.
 FT DOMAIN 663 667 CYTOSOLASMIC (POTENTIAL).
 FT CARBOHYD 591 591 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 630 630 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 667 AA: 68475 MW: 1DD95727032328BD CRC64;

Query Match 11.4%; Score 64.5; DB 1; Length 667;

Best Local Similarity 30.0%; Pred. No. 22;
 Matches 18; Conservative 7; Mismatches 18; Indels 17; Gaps 3;

OY 45 VLPDGS-SILRSYVALPTRD-----CLFEDPLHNGETMLKIMLNKFTGVIAFNC 94
 DB 89 VSPDGVCSVANEFYVPPNADATHDSVSCSEETPIH-----LANNKQYIGVACATC 141

RESULT 8

GLPE_HAEIN STANDARD; PRT: 105 AA.

ID GLPE_HAEIN
 AC P44819;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Thiosulfate sulfotransferase glpe (EC 2.8.1.1).
 GN GLPE OR HI0679.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7342800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton R., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fritchman J.L., Geoghagen N.S.M.,
 RA Guelm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
 RT Rd";
 RL Science 269:496-512(1995).

CC -1- FUNCTION: Catalyzes, although with low efficiency, the sulfur
 CC transfer reaction from thiosulfate to cyanide (By similarity).
 CC -1- CATALYTIC ACTIVITY: Thiosulfate + cyanide = sulfite + thiocyanate.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE GLPE FAMILY.
 CC -1- SIMILARITY: CONTAINS 1 RHODANESE DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC EMBL: U32750; AAC22338.1; -
 CC HSSP: P09390; IGMY.
 DR TIGR: HI0679; -
 DR InterPro: IPR001763; Rhodanese-like.
 DR Pfam: PF00581; Rhodanese; 1.
 DR SMART: SM00450; RHOD; 1.
 KW Transferase; Glycerol metabolism; Complete proteome.
 FT ACT SITE 63 63 BY SIMILARITY
 SO SEQUENCE 105 AA: 12033 MW: A3120E14BDD7245F CRC64;

Query Match 11.3%; Score 64; DB 1; Length 105;
 Best Local Similarity 21.9%; Pred. No. 3.2;
 Matches 25; Conservative 13; Mismatches 32; Indels 44; Gaps 4;

OY 3 WDMFGS-----THPCAFAHAAASRAISGPIYVSDSYGKINFDLKKLVLP 47
 DB 12 WEMMOGAGAILVDIRDMRPRAYSHPKGAFHJLNO-----SFLQFEELADF 55
 OY 48 DGSILRSEYVALPTRDCLFEDPLHNGETMLKIMLNKFTGVIGAFNCGGGGQWR 101
 DB 56 DSEPIVSCYHGVSSR-----NVATFLVEQGYKNVFSMTGGFD-----GWCRCR 96

```

RESULT 9
VN34_ROTPC STANDARD: PRT: 402 AA.
AC P27586;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Nonstructural RNA-binding protein 34 (NS34) (NCVp4).
GN S6.
OS Porcine rotavirus (group C / strain Cowden).
OC Viruses; dsRNA viruses; Reoviridae; Rotaviruses.
OX NCBI_TaxID=10916;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91361567; PubMed=1653496;
RA Qian Y.A., Jiang B.M., Sait L.J., Kang S.Y., Ojeh C.K., Green K.Y.,
RT "Molecular analysis of the gene 6 from a porcine group C rotavirus
RL that encodes the NS34 equivalent of group A rotaviruses.";
CC Virology 184:752-757(1991).
CC -1- SIMILARITY: CONTAINS 1 DBM (DOUBLE-STRANDED RNA-BINDING) DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: M69115; AAA47087.1; -.
DR PIR: A11040; MNXRPC.
DR InterPro: IPR001159; DS_RBD.
DR InterPro: IPR002873; Rota_NSP3.
DR Pfam: PF00035; dsrm: 1.
DR Pfam: PF01665; Rota_NSP3: 1.
DR PROSITE: PS50137; DS_RBD: 1.
KW Nonstructural protein: RNA-binding.
FT DOMAIN 384 400 DBM.
SQ SEQUENCE 402 AA: 45125 MW; 8B11F7AC7FB5135C CRC64;

Query Match 11.3%; Score 63.5; DB 1; Length 402;
Best Local Similarity 27.5%; Pred. No. 16;
Matches 28; Conservative 17; Mismatches 40; Indels 17; Gaps 5;

OY 8 STHPCAFAHAAASRAISGPIYVSDSVGKHNFD-----LLKKLYLPDGSILNSEYYA 58
DB 201 STNSTAAFYV---TIVGNPIKMYDESGKPLFDVGDYLNPKHIIDKMIENELPIFKSDYRN 257
OY 59 LPTRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGCGGMC 100
DB 258 NESPD--FD--VMNERSNLKIVSIND-CHALICVFKFENAMWC 294

RESULT 10
BBS2_RAT STANDARD: PRT: 721 AA.
AC 099MH9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Bardet-Biedl syndrome 2 protein homolog.
GN BBS2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21181710; PubMed=11285252;
RA Nishimura D.Y., Seardy C.C., Carmi R., Elbedour K., Van Maldergem L.,

```

```

RA Fulton A.B., Lam B.L., Powell B.R., Swiderski R.E., Bugge K.E.,
RA Haider N.B., Kwitek-Black A.E., Ying L., Duhl D.M., Gorman S.M.,
RA Heon E., Iannaccone A., Bonneau D., Biesecker L.G., Jacobson S.G.,
RA Stone E.M., Sheffield V.C.;
RT "Positional cloning of a novel gene on chromosome 16q causing
RT Bardet-Biedl syndrome (BBS2).";
RL Hum. Mol. Genet. 10:865-874(2001).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: AF342738; AAK28554.1; -.
DR SQ SEQUENCE 721 AA: 79917 MW; 55935DE19C0509DC CRC64;

Query Match 11.3%; Score 63.5; DB 1; Length 721;
Best Local Similarity 24.4%; Pred. No. 32;
Matches 21; Conservative 19; Mismatches 17; Indels 29; Gaps 6;

OY 6 FOSTHPCAFAHAAASRAISGPIYVSDSVGKHNFDLLKKLYLPDGSILNSEYYALPTRDCL 65
DB 24 YDGTHTPCLA--AATQA---GKVF-----HN-----PHMSQHPSTSR---V 57
OY 66 FEDPLHNGETMLKINLNKFTGVICA 91
DB 58 FOSPLESDVSL--NINOTVSCIGA 80

```

```

RESULT 11
Y083_NPVOP STANDARD: PRT: 819 AA.
AC 010336;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 91.1 kDa protein (ORF86).
OS Oryzias pseudoscutigata multicapsid polyhedrosis virus (OPMV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=164623;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97271300; PubMed=9126251;
RA Ahrens C.H., Russell R.R., Funk C.J., Evans J., Harwood S.,
RA Rohmann G.F.;
RT "The sequence of the Oryzias pseudoscutigata multinucleocapsid nuclear
RT polyhedrosis virus genome.";
RL Virology 229:381-399(1997).
CC -1- SIMILARITY: TO CORRESPONDING ORF IN ACNMPV.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: U75930; AAC59085.1; -.
DR InterPro: IPR002557; Chitin_bind_Pera.
DR Pfam: PF01607; CBM_14: 2.
DR SMART: SM00494; ChIBD2; 1.
KW Hypothetical protein.
SQ SEQUENCE 819 AA: 91067 MW; 108942AFOA8F925A CRC64;

Query Match 11.3%; Score 63.5; DB 1; Length 819;
Best Local Similarity 23.4%; Pred. No. 36;
Matches 25; Conservative 10; Mismatches 29; Indels 43; Gaps 5;

```

OY 1 PDWDFQ-----STHPCAAFAASRAISGPIPVSVSGAHNFDLKLKVLPGSILIR 53
 Db 261 PDQVFDRAIMNCVQTHHC-AFNGA-----CHYITIDIG----- 294
 OY 54 SEYVALPTRDCLFEDPLHNGEFTMLKIMLNKFTGYIGAFNCGCGMC 100
 Db 295 -----DAQFFKCLDNDREAOL-ITCINRWGADGQYACSGDARC 331
 RESULT 12
 TERT_YEAST STANDARD. PRT. 884 AA.
 AC Q06163;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Telomerase reverse transcriptase (EC 2.7.7.-) (Telomerase catalytic subunit).
 GN EST2 OR YLR318W OR L543.12.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota, Fungi, Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 OX NCBI_TaxID=4932;
 [1]
 SEQUENCE FROM N.A.
 RP STRAIN=S288c / AB972;
 RC MEDLINE=97313267; PubMed=9169871;
 RX Johnston M., Hillier L., Riles L., Albermann K., Andre B., Ansorge W., Barnes V., Bruckner M., Delius H., Dubois E., Dusterhoft A., Ertan K.D., Floeth M., Goffeau A., Goffeau A., Heblum K., Heuss-Neitzel D., Hilbert H., Hilger F., Kleine K., Kotter F., Louis E.J., Messenguy F., Mewes H.-W., Miosga T., Mostl D., Mueller-Auer S., Newtlich U., Obermayer B., Piravandi E., Pohl T.M., Portetlelle D., Purnelle B., Rechmann S., Rieger M., Rinke M., Rose M., Scharte M., Scherens B., Scholler P., Schweger C., Schwarz S., Underwood A.P., Uristarazu L.A., Vandenbol M., Verhasselt P., Vierendeels F., Voet M., Volckaert G., Voss H., Wambutt R., Wedler E., Wedler H., Zimmermann F.K., Zollner A., Hahl J., Honetzel U.D.;
 RA "The nucleotide sequence of Saccharomyces cerevisiae chromosome XII.";
 RT Nature 387:87-90(1997).
 RL [2]
 RP CHARACTERIZATION.
 RX MEDLINE=97274210; PubMed=9110970;
 RA Lingner J., Hughes T.R., Shevchenko A., Mann M., Lundblad V., Cech T.R.;
 RT "Reverse transcriptase motifs in the catalytic subunit of telomerase.";
 RL Science 276:561-567(1997).
 CC -I- FUNCTION: TELOMERASE IS A RIBONUCLEOPROTEIN ENZYME ESSENTIAL FOR THE REPLICATION OF CHROMOSOME TERMINI IN MOST EUKARYOTES. IT ELONGATES TELOMERES. IT IS A REVERSE TRANSCRIPTASE THAT ADDS SIMPLE SEQUENCE REPEATS TO CHROMOSOME ENDS BY COPYING A TEMPLATE SEQUENCE WITHIN THE RNA COMPONENT OF THE ENZYME.
 CC -I- SUBCELLULAR LOCATION: Nucleat.
 CC -I- MISCELLANEOUS: DELETION CAUSES TELOMERE SHORTENING AND SENESCENCE.
 CC -I- SIMILARITY: BELONGS TO THE REVERSE TRANSCRIPTASE FAMILY.
 CC TELOMERASE SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb.ch/announce/> or send an email to license@isb.ch).
 CC -----
 CC EMBL: U20618; AAB64520.1; -;
 CC SGD: S0004310; EST2;
 CC InterPro: IPR0004477; RYTse.
 CC InterPro: IPR003545; Telomerase_RT.
 CC Pfam: PF00078; tvL; 1.
 CC PRINTS: PRO1365; TELOMERASERT.
 CC -----

```

KW Transferrase; RNA-directed DNA polymerase; Telomere; Nuclear protein;
KW DNA-binding.
SQ SEQUENCE 884 AA; 102663 MW; 788334BA49592340 CRC64;

Query Match 11.28; Score 63; DB 1; Length 884;
Best Local Similarity 29.5%; Pred. No. 45;
Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4;

QY 26 PIYSDSGKINFDLKL--VLPDGSIL---RSEYVALPTDC--LFEEDPL--HNGET 75
   |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 499 PTQADRKRERKORLLKFNVPFLPELPMKDKVKSCTDISIPMECMRLKALKNENGF 558

QY 76 MUKTNLKNFTGVIGAFN 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 559 VRSQYFNTMTGVLKFN 576

RESULT 13
NIA_NEUCR ID NIA_NEUCR STANDARD; PRT; 982 AA.
AC P08619;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Nitrate reductase [NADPH] (EC 1.6.6.3) (NR).
GN NIT-3.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_Taxid=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=74-OR23-1A;
RX MEDLINE=91287699; PubMed=1829499;
RA Okamoto P.M., Fu Y.-H., Marzluf G.A.;
RT "Nit-3, the structural gene of nitrate reductase in Neurospora
RT crassa: nucleotide sequence and regulation of mRNA synthesis and
RT turnover.";
RL Mol. Gen. Genet. 227:213-223(1991).
RN [2]
RP PRELIMINARY PARTIAL SEQUENCE AROUND HIS-652.
RA Le K.H.D., Lederer F.;
RT "On the presence of a heme-binding domain homologous to cytochrome b5
RT in Neurospora crassa assimilatory nitrate reductase.";
RL EMBO J. 2:1909-1914(1983).
RN [3]
RP MUTANTS.
RC STRAIN=74-OR23-1A;
RX MEDLINE=93241176; PubMed=8479443;
RA Okamoto P.M., Garrett R.H., Marzluf G.A.;
RT "Molecular characterization of conventional and new repeat-induced
RT mutants of nit-3, the structural gene that encodes nitrate reductase
RT in Neurospora crassa.";
RL Mol. Gen. Genet. 238:81-90(1993).
RN [4]
RP MUTAGENESIS.
RX MEDLINE=93360901; PubMed=8355655;
RA Okamoto P.M., Marzluf G.A.;
RT "Nitrate reductase of Neurospora crassa: the functional role of
RT individual amino acids in the heme domain as examined by
RT site-directed mutagenesis.";
RL Mol. Gen. Genet. 240:221-230(1993).
CC -1- FUNCTION: NITRATE REDUCTASE IS A KEY ENZYME INVOLVED IN THE FIRST
CC STEP OF NITRATE ASSIMILATION IN PLANTS, FUNGI AND BACTERIA.
CC -1- CATALYTIC ACTIVITY: NADPH + nitrate = NADP(+) + nitrite + H(2)O.
CC -1- COFACTOR: EACH SUBUNIT CONTAINS 1 EQUIVALENT OF PAD, HEME IRON,
CC AND MOLYBDENUM-PTERIN 1 EQUIVALENT OF PAD, HEME IRON,
CC CALLED CYTOCHROME B-557.
CC -1- PATHWAY: NITRATE ASSIMILATORY PATHWAY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- INDUCTION: ITS EXPRESSION IS HIGHLY REGULATED AND RESPONDS RAPIDLY
CC TO NITRATE INDUCTION AND TO NITROGEN REPRESSION.
CC -1- SIMILARITY: TO EUKARYOTIC MOLYBDOPTERIN OXIDOREDUCTASES IN THE

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 24.6341 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 564

Sequence: 1 PDMDMFQSTHPCAFHAAASRAISGPIIYSDSGKINFDLKLVLPPDGSILRSEYALP 101

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PTCUTS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	564	100.0	784	US-08-846-234-5	Sequence 5, Appl
2	85	15.1	174	US-09-153-586-23	Sequence 23, Appl
3	63	11.2	884	US-08-851-843A-55	Sequence 55, Appl
4	63	11.2	884	US-08-974-549A-222	Sequence 222, App
5	63	11.2	884	US-08-854-050-55	Sequence 55, Appl
6	63	11.2	884	US-09-330-323-55	Sequence 55, Appl
7	61	10.8	3729	US-08-804-227C-4	Sequence 4, Appl
8	60.5	10.7	430	PCT-US93-04392-3	Sequence 3, Appl
9	60	10.6	210	US-08-948-616-5	Sequence 5, Appl
10	60	10.6	210	US-09-193-510-5	Sequence 5, Appl
11	60	10.6	210	US-09-368-402-5	Sequence 5, Appl
12	60	10.6	210	US-09-325-932A-179	Sequence 179, App
13	60	10.6	890	US-09-342-648-10	Sequence 10, Appl
14	59.5	10.5	430	PCT-US93-04392-6	Sequence 6, Appl
15	59.5	10.5	430	PCT-US93-04392-9	Sequence 9, Appl
16	59	10.5	390	US-09-586-719-2	Sequence 2, Appl
17	58.5	10.4	430	PCT-US93-04392-12	Sequence 12, Appl
18	58.5	10.4	430	PCT-US93-04392-15	Sequence 15, Appl
19	58.5	10.4	536	US-08-426-509A-12	Sequence 12, Appl
20	58.5	10.4	536	PCT-US95-05008-12	Sequence 12, Appl
21	58	10.3	210	US-08-948-616-10	Sequence 6, Appl
22	58	10.3	210	US-08-991-946A-6	Sequence 10, Appl
23	58	10.3	210	US-09-193-510-10	Sequence 10, Appl
24	58	10.3	210	US-09-368-402-10	Sequence 6, Appl
25	58	10.3	226	US-08-440-517A-6	Sequence 6, Appl
26	58	10.3	226	US-09-092-160-6	Sequence 11, Appl
27	58	10.3	228	US-08-766-982-11	Sequence 11, Appl

28	58	10.3	228	US-08-944-483-55	Sequence 55, Appl
29	58	10.3	228	US-09-296-219-11	Sequence 11, Appl
30	58	10.3	711	US-08-184-012C-8	Sequence 8, Appl
31	58	10.3	711	US-08-334-177-2	Sequence 2, Appl
32	58	10.3	711	US-08-666-082B-1	Sequence 1, Appl
33	58	10.3	711	US-08-766-982-2	Sequence 2, Appl
34	58	10.3	711	US-09-296-219-2	Sequence 2, Appl
35	58	10.3	711	PCT-US95-13830-2	Sequence 2, Appl
36	58	10.3	1584	US-09-251-645-6	Sequence 6, Appl
37	57.5	10.2	537	US-08-426-509A-11	Sequence 11, Appl
38	57.5	10.2	537	PCT-US95-05008-11	Sequence 11, Appl
39	57.5	10.2	954	US-09-057-969-2	Sequence 2, Appl
40	57	10.1	256	US-08-484-905-114	Sequence 114, App
41	57	10.1	256	US-08-481-985B-114	Sequence 114, App
42	57	10.1	256	US-08-370-476-114	Sequence 114, App
43	57	10.1	548	US-09-255-828-24	Sequence 24, Appl
44	57	10.1	858	US-09-255-829-22	Sequence 22, Appl
45	57	10.1	858	US-09-255-829-29	Sequence 29, Appl

ALIGNMENTS

```
RESULT 1
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-234-5

Query Match      100.0%   Score 564;   DB 4;   Length 784;
Best Local Similarity 100.0%;   Pred. NO. 1.2e-64;
Matches 101;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

OY      1 PDMDMFQSTHPCAFHAAASRAISGPIIYSDSGKINFDLKLVLPPDGSILRSEYALP 60
      |||
Db      510 PDMDMFQSTHPCAFHAAASRAISGPIIYSDSGKINFDLKLVLPPDGSILRSEYALP 569
OY      61 TRDCLFEDPLHNGETMLKLTWLNKFTGVIGAFNCGGCGMCR 101
      |||
```



```

: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/846,017
: FILING DATE: 25-APR-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/851,843
: FILING DATE: 06-MAY-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/854,050
: FILING DATE: 09-MAY-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/911,312
: FILING DATE: 14-AUG-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/912,951
: FILING DATE: 14-AUG-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/915,503
: FILING DATE: 14-AUG-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US97/17618
: FILING DATE: 01-OCT-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO PCT/US97/17885
: FILING DATE: 01-OCT-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph Ted
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 015389-002610US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 222:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 884 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: Protein
: LOCATION: 1..884
: OTHER INFORMATION: /note="Saccharomyces cerevisiae
: US-08-974-549A-222

Query Match          11.2% Score 63; DB 4; Length 884;
Best Local Similarity 29.5%; Pred. No. 24;
Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4;

QY 26 PIVSDSVGKHNFDLKLK-LVDPGSL- ---RSEYVALPTRDC--LPEDPL--HNGET 75
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 499 PTOIADRIKEFKORLLKFNNTVLPETYFMKFDVKSCYDSIPRMECRRLKDALKNNGFF 558
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 MLKIMNLNKFTGVIGAFN 93
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 559 VRSQYFNTNTGVLKLFN 576

RESULT 5
: US-08-854-050-55
: Sequence 55, Application US/08854050
: Patent No. 6261836
: GENERAL INFORMATION:
: APPLICANT: Cech, Thomas R.
: APPLICANT: Lingner, Joachim
: APPLICANT: Nakamura, Toru
: APPLICANT: Chapman, Karen B.
: APPLICANT: Morlin, Gregg B.
: APPLICANT: Harley, William H.
: APPLICANT: Andrews, William H.
: TITLE OF INVENTION: No. 6261836el Telomerase
: NUMBER OF SEQUENCES: 225
: CORRESPONDENCE ADDRESS:

```

```

: ADDRESSEE: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, 8th floor
: CITY: San Francisco
: STATE: California
: COUNTRY: United States of America
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/854,050
: FILING DATE: 09-MAY-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/851,843
: FILING DATE: 06-MAY-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/846,017
: FILING DATE: 25-APR-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/844,419
: FILING DATE: 18-APR-1997
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/724,643
: FILING DATE: 01-OCT-1996
: CLASSIFICATION: 536
: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph T.
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 015389-002930US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 884 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: US-08-854-050-55

Query Match          11.2% Score 63; DB 4; Length 884;
Best Local Similarity 29.5%; Pred. No. 24;
Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4;

QY 26 PIVSDSVGKHNFDLKLK-LVDPGSL- ---RSEYVALPTRDC--LPEDPL--HNGET 75
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 499 PTOIADRIKEFKORLLKFNNTVLPETYFMKFDVKSCYDSIPRMECRRLKDALKNNGFF 558
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 76 MLKIMNLNKFTGVIGAFN 93
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 559 VRSQYFNTNTGVLKLFN 576

RESULT 6
: US-09-430-323-55
: Sequence 55, Application US/09430323
: Patent No. 6309867
: GENERAL INFORMATION:
: APPLICANT: Cech, Thomas R.
: APPLICANT: Lingner, Joachim
: APPLICANT: Nakamura, Toru
: APPLICANT: Chapman, Karen B.
: APPLICANT: Morlin, Gregg B.
: APPLICANT: Harley, William H.
: APPLICANT: Andrews, William H.
: TITLE OF INVENTION: No. 6309867el Telomerase

```

```

1      NUMBER OF SEQUENCES: 225
2      CORRESPONDENCE ADDRESSES:
3      ADDRESSEE: Townsend and Townsend and Crew LLP
4      STREET: Two Embarcadero Center, 8th Floor
5      CITY: San Francisco
6      STATE: California
7      COUNTRY: United States of America
8      ZIP: 94111
9
10     COMPUTER READABLE FORM:
11     MEDIUM TYPE: Floppy disk
12     COMPUTER: IBM PC compatible
13     OPERATING SYSTEM: PC-DOS/MS-DOS
14     SOFTWARE: Patentin Release #1.0, Version #1.30
15     CURRENT APPLICATION DATA:
16     APPLICATION NUMBER: US/09/430,323
17     FILING DATE: 29-Oct-1999
18     CLASSIFICATION: <unknown>
19
20     PRIOR APPLICATION DATA:
21     APPLICATION NUMBER: US 08/854,050
22     FILING DATE: 09-MAY-1997
23     APPLICATION NUMBER: US 08/851,843
24     FILING DATE: 06-MAY-1997
25     APPLICATION NUMBER: US 08/846,017
26     FILING DATE: 25-APR-1997
27     APPLICATION NUMBER: US 08/844,419
28     FILING DATE: 18-APR-1997
29     APPLICATION NUMBER: US 08/724,643
30     FILING DATE: 01-OCT-1996
31     ATTORNEY/AGENT INFORMATION:
32     NAME: Apple, Randolph T.
33     REGISTRATION NUMBER: 36,429
34     REFERENCE/DOCKET NUMBER: 015389-00293005
35     TELECOMMUNICATION INFORMATION:
36     TELEPHONE: (415) 576-0200
37     TELEFAX: (415) 576-0300
38     INFORMATION FOR SEQ ID NO: 55:
39     SEQUENCE CHARACTERISTICS:
40     LENGTH: 884 amino acids
41     TYPE: amino acid
42     STRANDEDNESS: not relevant
43     TOPOLOGY: not relevant
44     MOLECULE TYPE: peptide
45     SEQUENCE DESCRIPTION: SEQ ID NO: 55:
46
47     US-09-430-323-55
48
49     Query Match      11.2%; Score 63; DB 4; Length 884;
50     Best Local Similarity 29.5%; Pred. No. 24;
51     Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4
52
53     Oy      26  PIYSDSVGKHNFDLTKL-LLPGSIL--RSEYVALPTRDC--LPEDDL--HNGFT 75
54             |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
55     Db      499 PQIADRIKEFQRLIKENNVLPFLYFAKFDVKSCIDSPMECKRIILKDALKNNGFF 558
56             ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
57     Oy      76  MKIWNLKFTGVIGAFN 93
58             ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
59     Db      559 VRSQYFENTMGVLKFN 576
60
61     RESULT 7
62     US-08-804-227C-4
63     Sequence 4, Application US/08804227C
64     Patent No. 5876991
65     GENERAL INFORMATION:
66     APPLICANT: Dehoff, Bradley S.
67     APPLICANT: Kuhstoss, Stuart A.
68     APPLICANT: Kosteck, Paul R., Jr.
69     APPLICANT: Sutton, Kimberly L.
70     TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
71     NUMBER OF SEQUENCES: 15
72     CORRESPONDENCE ADDRESSES:
73     ADDRESSEE: THOMAS G. PLANT 1501
74     STREET: LULUY CORPORATE CENTER
75     CITY: INDIANAPOLIS

```

```

1 STATE: IN
2 COUNTRY: USA
3 ZIP: 46285
4
5 COMPUTER READABLE FORM:
6 MEDIUM TYPE: Floppy disk
7 COMPUTER: IBM Compatible
8 OPERATING SYSTEM: MS-DOS
9 SOFTWARE: ASCII(DOS) Text only
10
11 CURRENT APPLICATION DATA:
12 APPLICATION NUMBER: US/08/804,227C
13 FILING DATE: February 21, 1997
14
15 CLASSIFICATION: 435
16
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Plant, Thomas, G.
19 REGISTRATION NUMBER: 35,784
20 REFERENCE/DOCKET NUMBER: X-8231
21
22 TELECOMMUNICATION INFORMATION:
23 TELEPHONE: 317-276-2459
24
25 INFORMATION FOR SEQ ID NO: 4:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 3729 amino acids
28 TYPE: amino acid
29 TOPOLOGY: unknown
30
31 MOLECULE TYPE: protein
32
33 US-08-804-227C-4
34
35 Query Match 10.8%; Score 61; DB 2; Length 3729;
36 Best Local Similarity 30.9%; Pred. NO. 3.2e+02;
37 Matches 17; Conservative 13; Mismatches 21; Indels 4; Gaps 3
38
39
40 Oy 8 STHPCAATHA--ASRAISGGPIYVSDSVG-KINFDLKKLVLPDGSILRSEYAL 59
41 :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
42 Db 2602 TTDP-AAFYAEFAERGGYDYGPAFOGFTAGARGEDVVAEVALPSGLVADARRHRL 2655
43
44 RESULT 8
45 PCT-US93-04392-3
46
47 Sequence 3, Application PC/TUS9304392
48
49 GENERAL INFORMATION:
50
51 APPLICANT: Beck, Steven R., Cain, Robert O., Chan, Hardy W.
52
53 APPLICANT: Freedman, Richard, Heefner, Donald L.
54
55 APPLICANT: Phelephs, Trish, Roberts, Christopher R.
56
57 APPLICANT: Salazar, Felix H., Snyder, Roger C.
58
59 TITLE OF INVENTION: Enzymatic Process for Production of
60
61 NUMBER OF SEQUENCES: 15
62
63 CORRESPONDENCE ADDRESS:
64
65 ADDRESSSEE:
66
67 STREET:
68
69 CITY:
70
71 STATE:
72
73 COUNTRY:
74
75 ZIP:
76
77 COMPUTER READABLE FORM:
78
79 MEDIUM TYPE: Floppy disk
80
81 COMPUTER: IBM PC compatible
82
83 OPERATING SYSTEM: PC-DOS/MS-DOS
84
85 SOFTWARE: PatentIn Release #1.0, Version #1.25
86
87 CURRENT APPLICATION DATA:
88 APPLICATION NUMBER: PCT/US93/04392
89 FILING DATE: 19930514
90
91 CLASSIFICATION:
92
93 PRIOR APPLICATION DATA:
94
95 APPLICATION NUMBER: US/07/883,658
96 FILING DATE: MAY 15, 1992
97
98 INFORMATION FOR SEQ ID NO: 3:
99
100 SEQUENCE CHARACTERISTICS:
101 LENGTH: 430 amino acids
102 TYPE: AMINO ACID
103 STRANDEDNESS: double
104 TOPOLOGY: circular
105
106 MOLECULE TYPE: cDNA
107
108 HYPOTHETICAL: YES
109
110 ANTI-SENSE: NO

```

; ORIGINAL SOURCE:
 ; ORGANISM: Zopfiella latipes
 ; STRAIN: 511
 ; INDIVIDUAL ISOLATE: ATCC #26183
 ; IMMEDIATE SOURCE:
 ; CLONE: CL511 1-2
 ;
 PCT-US93-04392-3

Query Match	10.7%	Score 60.5;	DB 5;	Length 430;
Best Local Similarity	25.5%	Pred. No. 18;		
Matches 25; Conservative	15;	Mismatches 39;	Indels 19;	Gaps 4;

Qy 10 HPCAFHAASRAISG--PIYSD-----SVGHNDLKKVLPPDGSILRSEIYA 58
|| : ::| :| : :: | :| :| :|
Db 248 HPDDLATGKQVLAGCGNMNLVADGDFGSHGMFTTGENYLKVLKSLANDGKLSPKVYN 307

```

Oy      59  LPTRDCLFEDPLHNGETMLKIWNLNKKTGVIGAFNCQG  96
          |   |   |   |   |   |   |   |   |   |
Db      308  L-----MFEHLTGG---AKKGHEDLNLPVSGFFAVG  337

```

RESULT 9
US-08-948-616-5

```

: GENERAL INFORMATION:
:
: APPLICANT: HILLMAN, Jennifer L.
:
: APPLICANT: Lal, Preeti
:
: APPLICANT: Shah, Purni
:
: APPLICANT: Corley, Neil C.
:
: TITLE OF INVENTION: VESTICLE TRANSPORT ASSOCIATED PROTEINS
:
: NUMBER OF SEQUENCES: 11
:
: CORRESPONDENCE ADDRESS:
:

```

```

1 COMPUTER READABLE FORM:
2 MEDIUM TYPE: Diskette
3 COMPUTER: IBM Compatible
4 OPERATING SYSTEM: DOS
5 SOFTWARE: FASTEST for Windows Version 2.0
6
7 CURRENT APPLICATION DATA:
8 APPLICATION NUMBER: US/08/948,616
9 FILING DATE: Herewith
10 CLASSIFICATION: 530
11 PRIOR APPLICATION DATA:
12 APPLICATION NUMBER:
13

```

CLONE: 2620104
US-08-948-616-5

Query Match	10.6%	Score 60;	DB 2;	Length 210;
Best Local Similarity	33.3%	Pred. No. 7.8;		
Matches	11;	Conservative	10;	Mismatches 10;
			Indels	2;
			Gaps	1.

```
QY 22 ISGGPI--YVSDSVGKHNEDLLKLVLPDGSIL 522
      : | | : : | : : | | | |
Db 60 VKGKPVHLHIWDTAGQDDYDRLRPLFYPDASVL 922
```

RESULT 10
US-09-193-510-5

? GENERAL INFORMATION:
 ? APPLICANT: Hillman, Jennifer L.
 ? APPLICANT: Lal, Preeti
 ? APPLICANT: Shah, Preeti
 ? APPLICANT: Corley, Neil C.
 ? TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
 ? NUMBER OF SEQUENCES: 11
 ? CORRESPONDENCE ADDRESS: 11
 ?

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Diskette
3  COMPUTER: IBM Compatible
4  OPERATING SYSTEM: DOS
5  SOFTWARE: FASTSO for Windows Version 2.0
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US/09/193.510
8  FILING DATE:

```

CLONE: 2620104
US-09-193-510-5

Query Match	10.6%	Score 60	DB 2	Length 210
Best Local Similarity	33.3%	Pred. No. 7.8		
Matches 11	Conservative	10	Indels	2
			Gaps	1

```

0y      22 ISGPI--YVSDSVGKHNFDLKLVLPGDSIL 52
          : | | : : | | : : | | : : | | : : |
Db      60 VKGKPVHLHWTAGQDDYDRLRLPEFYPDASVL 92

```

RESULT 11

US-09-366-402-5
; Sequence 5, Application US/09366402
; Patent No. 6200790
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.

```

; TITLE OF INVENTION: VESICLE TRANSPORT ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/368,402
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/193,510
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0409 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 210 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: KERANOT02
; CLONE: 2620104
;
; US-09-368-402-5
;
Query Match 10.6%; Score 60; DB 4; Length 210;
Best Local Similarity 33.3%; Pred. No. 7.8;
Matches 11; Conservative 10; Mismatches 10; Indels 2; Gaps 1;

QY 22 ISGGPI--YVSDSVGKHNFDLKLKLVLPDGSIL 52
;| | : : | | : : | | : : | | : : |
Db 60 VKGKPVHLHIMDTAGDDYDRLRLPFLYPDASVL 92

RESULT 12
; US-09-325-932A-179
; Sequence 179, Application US/09325932A
; Patent No. 6451604
; GENERAL INFORMATION:
; APPLICANT: Flinn, Barry
; APPLICANT: Lasham, Annette
; TITLE OF INVENTION: Compositions affecting programmed cell
; TITLE OF INVENTION: death and their use in the modification of forestry plant develo
; FILE REFERENCE: 1022
; CURRENT APPLICATION NUMBER: US/09/325,932A
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 206
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 179
; LENGTH: 608
; TYPE: PRT
; ORGANISM: Pinus radiata
;
; US-09-325-932A-179
;
Query Match 10.6%; Score 60; DB 4; Length 608;
Best Local Similarity 21.4%; Pred. No. 34;
Matches 25; Conservative 18; Mismatches 48; Indels 26; Gaps 5;
```

```

QY 3 WDMFOSTHPCAFHAA-----SRAISGPIYVSDSVGKHNFDLK-KLVLPD 48
;| | : : | | : : | | : : | | : : |
Db 427 WHAFGRSPVVGFEKYEKFLNACNGLPILSKVLGALLHCKDDKLKNAQLRRTSKVLE 486

QY 49 G--SIRSEYVALPTRD-----CLFEDPLHNGETMLKIWNINKTYIGAFNCQ 95
;| | : : | | : : | | : : | | : : |
Db 487 DIRSLRISYDALDKERQIFLDIACFFIG--KNRDSAIRVWDSNMEGLGLMKLE 541

RESULT 13
; US-09-342-648-10
; Sequence 10, Application US/09342648
; Patent No. 6248584
; GENERAL INFORMATION:
; APPLICANT: Cahoon, Rebecca E.
; APPLICANT: Odell, Joan
; APPLICANT: Rafalski, Antoni
; TITLE OF INVENTION: Transcription Coactivators
; FILE REFERENCE: BB-1169-B
; CURRENT APPLICATION NUMBER: US/09/342,648
; CURRENT FILING DATE: 1999-06-29
; EARLIER APPLICATION NUMBER: 60/092,659
; EARLIER FILING DATE: July 13, 1998
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 10
; LENGTH: 890
; TYPE: PRT
; ORGANISM: Ajellomyces capsulatus
;
; US-09-342-648-10
;
Query Match 10.6%; Score 60; DB 4; Length 890;
Best Local Similarity 25.0%; Pred. No. 59;
Matches 19; Conservative 9; Mismatches 22; Indels 26; Gaps 3;

QY 48 DGSILSEYVALPTRDCLFED-----PLHNGETML-KIWNINKT 85
;| | : : | | : : | | : : | | : : |
Db 39 DGSILSYYPALTAOPAFQSRFLRELLGVKVFQVDYVPTVRNGNMIVLHNQEN 98

QY 86 TGVIQAFNCGGCMCR 101
;| | : : | | : : | | : : |
Db 99 LAEL-----CYAEGWVK 110

RESULT 14
; PCT-US93-04392-6
; Sequence 6, Application PC/TUS9304392
; GENERAL INFORMATION:
; APPLICANT: Beck, Steven R., Cain, Robert O., Chan, Hardy W.
; APPLICANT: Freedman, Richard, Heefner, Donald L.
; APPLICANT: Phelps, Trish, Roberts, Christopher R.
; APPLICANT: Salazar, Felix H., Snyder, Roger C.
; TITLE OF INVENTION: Enzymatic Process for Production of
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04392
; FILING DATE: 19930514
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/883,658
; FILING DATE: MAY 15, 1992
```

```
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Zopfiella latipes
; STRAIN: 780
; INDIVIDUAL ISOLATE: ATCC #44575
; IMMEDIATE SOURCE:
; CLONE: 21780-3A
PCT-US93-04392-6

Query Match 10.5%; Score 59.5; DB 5; Length 430;
Best Local Similarity 25.5%; Pred. No. 25;
Matches 24; Conservative 15; Mismatches 36; Indels 19; Gaps 4;

Qy 10 HPCAFHNASRAISGG--PIYSD-----SVGKHNFLLKKLYLPDGSILRSEYYA 58
Db 248 HPDDPLATGKQVLAGGNNLVADGDFGCHGMFTTGENTLKVLSLANDGKLSPEMYN 307

Qy 59 LPTRDCLFEDPLHNGETMLKINLNKFTGVIGAF 92
Db 308 L-----MFEHDLTGG---AKKGHEDALNGPVGSF 333

RESULT 15
PCT-US93-04392-9
; Sequence 9, Application PC/TUS9304392
; GENERAL INFORMATION:
; APPLICANT: Beck, Steven R., Cain, Robert O., Chan, Hardy W.
; APPLICANT: Freedman, Richard, Heefner, Donald L.
; APPLICANT: Pnelphs, Trish, Roberts, Christopher R.
; APPLICANT: Salazar, Felix H., Snyder, Roger C.
; TITLE OF INVENTION: Enzymatic Process for Production of
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE:
; STREET:
; CITY:
; STATE:
; COUNTRY:
; ZIP:
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04392
; FILING DATE: 19930514
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/883,658
; FILING DATE: MAY 15, 1992
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 430 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Zopfiella latipes
; STRAIN: 780
; INDIVIDUAL ISOLATE: ATCC #44575
; IMMEDIATE SOURCE:
```

```
; LIBRARY: CDNA in lambda-gtl1
; CLONE: 21780-m10
PCT-US93-04392-9

Query Match 10.5%; Score 59.5; DB 5; Length 430;
Best Local Similarity 25.5%; Pred. No. 25;
Matches 24; Conservative 15; Mismatches 36; Indels 19; Gaps 4;

Qy 10 HPCAFHNASRAISGG--PIYSD-----SVGKHNFLLKKLYLPDGSILRSEYYA 58
Db 248 HPDDPLATGKQVLAGGNNLVADGDFGCHGMFTTGENTLKVLSLANDGKLSPEMYN 307

Qy 59 LPTRDCLFEDPLHNGETMLKINLNKFTGVIGAF 92
Db 308 L-----MFEHDLTGG---AKKGHEDALNGPVGSF 333
```

Search completed: March 27, 2003, 10:07:36
Job time : 25.6341 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 71.439 Seconds
(without alignments)
188.389 Million cell updates/sec

Title: US-09-675-208-5_COPY_510_610

Perfect score: 564
Sequence: 1 PDMDFFGTHPCAFHNAASR.....LNKFTGVIGAFNCQGGWCR 101

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A.Geneseq_101002:*
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	564	100.0	784	19	AAW53570
2	564	100.0	784	20	AAV17417
3	515	91.3	587	19	AAW57888
4	515	91.3	587	20	AAV30144
5	511	90.6	783	23	ABB93664
6	508	90.1	572	20	AAV32075
7	506	89.7	777	20	AAV32074
8	505	89.5	799	19	AAW57886
9	505	89.5	799	20	AAV30142
10	498	88.3	780	22	AAW98659

11	497	88.1	781	19	AAW57887
12	497	88.1	781	20	AAV30143
13	497	88.1	781	22	AAW49400
14	494	87.6	758	21	AAV70978
15	476	84.4	783	20	AAV32073
16	444	78.7	265	20	AAV32072
17	426	75.5	271	19	AAW57889
18	426	75.5	271	20	AAV30145
19	404	71.6	751	21	AAV70980
20	403	71.5	756	21	AAV70975
21	400	70.9	763	21	AAV70977
22	394	69.9	508	21	AAV70974
23	387	68.6	750	20	AAV17418
24	387	68.6	770	21	AAV70976
25	384	68.1	841	21	AAV70981
26	371	65.8	807	23	ABB92806
27	205	36.3	229	21	AAW28628
28	85	15.1	174	20	AAV06803
29	79	14.0	15	20	AAV17421
30	79	14.0	193	21	AAW28629
31	71	12.6	638	22	ABB64380
32	69.5	12.3	412	18	AAW20443
33	69	12.2	426	22	AAW90870
34	69	12.2	615	22	AAU28187
35	69	12.2	615	22	AAW93480
36	68.5	12.1	183	22	AAU33604
37	68.5	12.1	466	22	ABG09425
38	68.5	12.1	466	22	ABG27212
39	67	11.9	1778	22	AAW79480
40	67	11.9	2099	22	AAW78496
41	65.5	11.6	336	23	AAW92001
42	64	11.3	359	23	ABB93733
43	63.5	11.3	207	18	AAW10493
44	63.5	11.3	423	23	ABP27573
45	63.5	11.3	625	20	AAV31978

ALIGNMENTS

RESULT 1	
AAW53570	
ID	AAW53570 standard; Protein: 784 AA.
XX	
AC	AAW53570;
ID	06-JUL-1998 (first entry)
DT	
XX	
DE	Cucurbit raffinose synthase.
XX	
KW	Cucurbit; raffinose synthase; sucrose; galactinol.
XX	
OS	Cucumis sativus.
XX	
PN	JP10084973-A.
XX	
PD	07-APR-1998.
XX	
PF	28-APR-1997; 97JP-0111124.
XX	
PR	26-JUL-1996; 96JP-0198079.
PR	26-APR-1996; 96JP-0107682.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
DR	WPI: 1998-264858/24.
DR	N-PSDB: AAV22250.
XX	
PT	Raffinose synthase gene - useful for preparation of raffinose in
PT	transformed plant
XX	
PS	Claim 3; Pages 17-20; 26pp; Japanese.
XX	

Soybean raffinose
Amino acid sequenc
Soybean raffinose
Soybean raffinose
Soybean raffinose
Sugarteet raffinose
Soybean raffinose
Corn raffinose syn
Amino acid sequenc
Wheat raffinose sy
Corn raffinose syn
Rice raffinose syn
Corn raffinose syn
Soybean raffinose
Rice raffinose syn
Wheat raffinose sy
Herbicideally activ
Arabidopsis thalia
Peptide Seq ID No:
Soybean raffinose
Arabidopsis thalia
Drosophila melanog
H. pylori cytoplas
C glutamicum prote
Novel human secret
Human protein sequ
Pseudomonas aerugi
Novel human diagno
Human protein SGO
Human protein SGO
Herbicideally activ
Soluble fused MHC
Streptococcus poly
Mouse melanoma ant

CC The present sequence is cucumber raffinose synthase, which
CC forms raffinose from sucrose and galactinol, has an optimum pH of
CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
CC lodosacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA;

Query Match 100.0%; Score 564; DB 19; Length 784;
Best Local Similarity 100.0%; Pred. No. 3.3e-63;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHFLLKRLVLPDGSILRSEYALP 60

DB 510 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHFLLKRLVLPDGSILRSEYALP 569

OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 101

DB 570 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 610

RESULT 2

AAV17417

ID AAV17417 standard; Protein; 784 AA.

XX AC AAV17417;

DT 29-JUL-1999 (first entry)

XX DE Cucumber raffinose synthase.

XX KW Raffinose synthase; sucrose; galactinol.

XX OS Cucumis sativus.

XX PN JPI1123080-A.

XX PD 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

PA (AJIN) AJINOMOTO KK.

DR WPI: 1999-340516/29.

DR N-PSDB; AAV61238.

PT New raffinose synthase gene - for production of raffinose from
PT sucrose and galactinol

PS Claim 2: Page 25-27; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from cucumber.

XX SQ Sequence 784 AA;

Query Match 100.0%; Score 564; DB 20; Length 784;

Best Local Similarity 100.0%; Pred. No. 3.3e-63;
Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHFLLKRLVLPDGSILRSEYALP 60

DB 510 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHFLLKRLVLPDGSILRSEYALP 569

OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 101

DB 570 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 610

RESULT 3

AAW57888

ID AAW57888 standard; Protein; 587 AA.

XX AC AAW57888;

DT 23-SEP-1998 (first entry)

XX DE Japanese artichoke raffinose synthetase.

XX KW Raffinose synthetase; metabolism modification; food additive;

XX KM gastrointestinal flora; Japanese artichoke.

XX OS Stachys sieboldii.

XX PN EP849359-A2.

PD 24-JUN-1998.

PF 18-DEC-1997; 97EP-0122417.

PR 18-DEC-1996; 96JP-0338673.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Oeda K, Mantanabe E;

DR WPI: 1998-324670/29.

DR N-PSDB; AAV40802.

PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora

XX Claim 1: Page 36-38; 44pp; English.

CC This sequence is the Japanese artichoke raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.

XX SQ Sequence 587 AA;

Query Match 91.3%; Score 515; DB 19; Length 587;
Best Local Similarity 89.1%; Pred. No. 4.4e-57;
Matches 90; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

OY 1 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHFLLKRLVLPDGSILRSEYALP 60

DB 401 PDMDMFQSTHPCAFHAAASRAISGPIYVSDVSGKHFLLRSLVLPDGSILRSEYALP 460

OY 61 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 101

DB 461 TRDCLFEDPLHNGETMLKINLNKFTGVIGAFNCGGGWCGR 501

RESULT 4

AAV30144

ID AAV30144 standard; Protein; 587 AA.

XX AC AAV30144;

DT 26-OCT-1999 (first entry)

XX Amino acid sequence of a raffinose synthase protein.

XX	Raffinose synthase; plant; sucrose; raffinose.
XX	
OS	Stachys sieboldii.
XX	
PN	JF11215984-A.
XX	
PD	10-AUG-1999.
XX	
PF	12-DEC-1997; 97JP-0342899.
XX	
PR	28-NOV-1997; 97JP-0329006.
PR	18-DEC-1996; 96JP-0338673.
XX	
PA	(SUMO) SUMITOMO CHEM CO LTD.
XX	
DR	WPI: 1999-511112/43.
DR	N-PSDB: AA210003.
XX	
PT	New raffinose synthase gene - is prepared from a plant material
XX	
PS	Claim 12; Page 30-31; 40pp; Japanese.
XX	
CC	The present sequence represents a raffinose synthase protein. The
CC	sequence is isolated from plant material. The protein forms raffinose
CC	by complexing alpha(1 to 6) - D-galactosyl hydroxyl group of the 6C
CC	of D-glucose residue in sucrose molecules.
XX	
XX	
Sequence	587 AA;

Query Match	91.3%	Score 515	DB 20	Length 587
Best Local Similarity	89.1%	Pred. No. 4.4e-57		
Matches 90; Conservative	5;	Mismatches 6;	Indels 0;	Gaps 0

QY 1 PDMMFQSTHPCAAFFHAAASRIISGCPPIYSDSYGKHNHFDKKIKYLPDGLSEAYALP 60
Db 401 PDMMFQSTHPCAAFFHAAASRIISGCPPIYSDSYGKHNHFDKKIKYLPDGLSEAYALP 460
QY 61 TRDCLFEDPLHNGETMLKIMNLNKFVIGAFINQGGGMCGR 101
Db 461 TRDCLFEDPLHNGETMLKIMNLNKFVIGAFINQGGGMCGR 501

RESULT 5
ABB93664
ID ABB93664 standard; Protein; 783 AA

AC ABB93664;

DT 31-MAY-2002 (first entry)

DE Herbicidally active polypeptide SEQ ID NO 2875.

KW Herbicidal; plant; agriculture; herbicide

OS *Arabidopsis thaliana*.

PN WO200210210-A2.

PD 07-FEB-2002.

PF 28-AUG-2001; 2001WO-EP09892.

PR 28-AUG-2001; 2001WO-EP09892.

PA (FARB) BAYER AG.

PI Tietjen K, Weldler M;

DR WPI; 2002-269010/31.

PT Identifying plant target proteins for herbicidally active compounds, PT comprising aligning and comparing nucleic acid or amino acid sequences

PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 2875; 261pp + Sequence Listing; English.

CC The invention relates to identifying target proteins
CC (ABR90790-ABR94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.

SQ Sequence 783 AA;

Query Match	90.68;	Score 511;	DB 23;	Length 783;
Best Local Similarly	89.18;	Pred. No. 2.2e-56;		
Matches 90; Conservative	7;	Mismatches 4;	Indels 0;	Gaps 0;

Qy	Db	Qy
1	1	61
PDMDPFQSTHPCAAEFHMA	PDMDPFQSTHPCAAEFHMA	TRDCLFEDPLHNETMLK
RAISGPIYVSQKHNKDLK	RAISGPIYVSQKHNKDLK	INLNKLFYIGAFNCGG
VLDPDSTIREYALP	VLDPDSTIREYALP	GMCR 101
60	60	
515	515	575
PDMDPFQSTHPCAAEFHMA	PDMDPFQSTHPCAAEFHMA	TRDRLFEEDPLHDGKTM
RAISGPIYVSQKHNKDLK	RAISGPIYVSQKHNKDLK	LNKLFYIGAFNCGG
VLDPDSTIREYALP	VLDPDSTIREYALP	GMCR 615
574	574	

RESULT 6
 AAY32075
 ID AAY32075 standard; Protein; 572 AA

AC AAY32075;

DT 17-JAN-2000 (first entry)

DE Rapeseed raffinose synthase.

KW Raffinose synthase; rapeseed; transgenic plant.

OS Brassica napus.

FH	Key	Location/Qualifiers

/note= "encoded by GGY"

FT /note- "encoded by GW"

/note="encoded by CCS"

FT /note="encoded by TCR"

/note= "encoded by CGR"

PN EP953643-A2.

PD 03-NOV-1999

PF 27-APR-1999; 99EP-0107430.

PR 30-APR-1998; 98JP-0120550.

PR 04-DEC-1998; 98JP-0345590.

XX
XX
XX

24 XX

DR WPI; 1999-593144/51.

DR N-PSDB; AAZ20210.

XX New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX
XX
PS Claim 27, Page 36-38; 55pp; English.
XX
CC This sequence represents rapessed raffinose synthase, a protein
CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC hydroxy group attached to the carbon atom at the 6-position of the
CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
CC (see AA20210) encoding the enzyme was isolated from rapessed cv.
CC Westar leaf cDNA by PCR. Probes or primers generated from plant
CC raffinose synthase genes (see AA20207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
XX
SQ Sequence 572 AA;

Query Match 90.1%; Score 508; DB 20; Length 572;
Best Local Similarity 87.1%; Pred. No. 3.4e-56;
Matches 88; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHNASRAISGPIYSDSVGNHFDLKLKVLPGSILRSEYALP 60
DB 304 PDMDMFQSTHPCAFHNASRAISGPIYSDCVQHDFDLRLRVLPDGSILRCEYALP 363
DB 61 TRDCLFEDPLHNGETMLKINLNKFTGVICAFNCGGWCRCR 101
DB 364 TRDRLEFDPLHDGKTMKINLNKFTGIIIGAFNCGGWCRCR 404

RESULT 7
AAI32074
ID AAY32074 standard; Protein; 777 AA.
AC AAY32074;
XX
DT 17-JAN-2000 (first entry)
XX
DE Mustard raffinose synthase.
XX
KW Raffinose synthase; mustard; transgenic plant.
XX
OS Brassica juncea.
XX
FH Key Location/Qualifiers
FT Misc-difference 210 /note= "encoded by ACR"
PT
PN EP953643-A2.
XX
PD 03-NOV-1999.
XX
PE 27-APR-1999; 99EP-0107430.
XX
PR 30-APR-1998; 98JP-0120550.
PR 30-APR-1998; 98JP-0120551.
PR 04-DEC-1998; 98JP-0345590.
PR 10-DEC-1998; 98JP-0351246.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Watanabe E, Oeda K;
XX
DR WPI; 1999-593144/51.
DR N-PSDB; AA220209.
XX
PT New sense and antisense genes, useful for altering the level of

PT raffinose in food plants -
XX
XX
PS Claim 26, Page 29-31; 55pp; English.
XX
CC This sequence represents mustard raffinose synthase, a protein
CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC hydroxy group attached to the carbon atom at the 6-position of the
CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
CC (see AA20209) encoding the enzyme was isolated from mustard
CC (Brassica juncea) leaf cDNA by PCR. Probes or primers generated from
CC plant raffinose synthase genes (see AA20207-10) may be used to obtain
CC other raffinose synthase genes by labeled detection or amplification
CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.
XX
SQ Sequence 777 AA;

Query Match 89.7%; Score 506; DB 20; Length 777;
Best Local Similarity 87.1%; Pred. No. 9.4e-56;
Matches 88; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHNASRAISGPIYSDSVGNHFDLKLKVLPGSILRSEYALP 60
DB 509 PDMDMFQSTHPCAFHNASRAISGPIYSDCVQHDFDLRLRVLPDGSILRCEYALP 568
QY 61 TRDCLFEDPLHNGETMLKINLNKFTGVICAFNCGGWCRCR 101
DB 569 TRDRLEFDPLHDGKTMKINLNKFTGIIIGAFNCGGWCRCR 609

RESULT 8
AAW57886
ID AAW57886 standard; Protein; 799 AA.
AC AAW57886;
XX
DT 23-SEP-1998 (first entry)
XX
DE Broad bean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; broad bean.
XX
OS Vicia faba.
XX
FN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PE 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Watanabe E;
XX
DR WPI; 1998-324670/29.
DR N-PSDB; AAV40800.
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
PS Claim 1, Page 26-29; 44pp; English.
XX
CC This sequence represents the broad bean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a

CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.

XX Sequence 799 AA:

Query Match 89.5%; Score 505; DB 19; Length 799;

Best Local Similarity 88.0%; Pred. No. 1.3e-55;

Matches 88; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 PDMDFQSTHPCAFHAAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 60

DB 529 PDMDFQSTHPCAFHAAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 588

OY 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMC 100

DB 589 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMC 628

RESULT 9

AAV30142

XX AAV30142 standard; Protein; 799 AA.

AC AAV30142;

DT 26-OCT-1999 (first entry)

DE Amino acid sequence of a raffinose synthase protein.

XX Raffinose synthase; plant; broad bean; sucrose; raffinose.

OS Vicia faba.

PN JP11215984-A.

PD 10-AUG-1999.

PF 12-DEC-1997; 97JP-0342899.

PR 28-NOV-1997; 97JP-0329006.

PR 18-DEC-1996; 96JP-0338673.

PA (SUMO) SUMITOMO CHEM CO LTD.

DR WPI; 1999-511112/43.

DR N-PSDB; AA210001.

PT New raffinose synthase gene - 1s prepared from a plant material

PS Claim 5; Page 19-21; 40pp; Japanese.

CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material of broad beans. The
CC protein forms raffinose by complexing alpha(1 to 6) - D-galactosyl
CC hydroxyl group of the 6C of D-glucose residue in sucrose molecules.

XX Sequence 799 AA:

Query Match 89.5%; Score 505; DB 20; Length 799;

Best Local Similarity 88.0%; Pred. No. 1.3e-55;

Matches 88; Conservative 5; Mismatches 7; Indels 0; Gaps 0;

OY 1 PDMDFQSTHPCAFHAAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 60

DB 529 PDMDFQSTHPCAFHAAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 588

OY 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMC 100

DB 589 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMC 628

RESULT 10

AAB98659

ID AAB98659 standard; protein; 780 AA.

XX AAB98659;

DT 17-AUG-2001 (first entry)

DE Soybean protein; SEQ ID 1.

XX Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;

KW plant; soybean.

OS Glycine max.

PN JP2001078783-A.

PD 27-MAR-2001.

PF 03-JUL-2000; 2000JP-0200571.

PR 09-JUL-1999; 99JP-0196036.

PA (SUMO) SUMITOMO CHEM CO LTD.

DR WPI; 2001-313373/33.

DR N-PSDB; AAH27438.

PT Novel mutant protein of raffinose synthase is useful for reducing the
PT raffinose oligosaccharide content in a plant body -

PS Disclosure; Page 18-20; 30pp; Japanese.

CC The present invention relates to a mutant protein of raffinose synthase
CC in which at least one aromatic amino acid present at the position of
CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC mutant protein can be used for reducing the raffinose oligosaccharide
CC content in a plant body. The present protein from soybean, was used in
CC the present invention.

XX Sequence 780 AA:

Query Match 88.3%; Score 498; DB 22; Length 780;

Best Local Similarity 86.0%; Pred. No. 1e-54; Mismatches 7; Indels 0; Gaps 0;

Matches 86; Conservative 7; Mismatches 7; Indels 0; Gaps 0;

OY 1 PDMDFQSTHPCAFHAAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 60

DB 510 PDMDFQSTHPCAFHAAASRAISGPIYVSDVGKHNFDLKKVLPDGSILRSEYALP 569

OY 61 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMC 100

DB 570 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGGMC 609

RESULT 11

ID AAV57887 standard; Protein; 781 AA.

XX AAV57887;

DT 23-SEP-1998 (first entry)

DE Soybean raffinose synthetase.

XX Raffinose synthetase; metabolism modification; food additive;

KW gastrointestinal flora; soybean.

OS Glycine max.

PN EP849359-A2.

```

PD 24-JUN-1998.
XX
XX 18-DEC-1997; 97EP-0122417.
XX
XX 18-DEC-1996; 96JP-0338673.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Oeda K, Watanabe E;
XX
XX MPI: 1998-324670/29.
XX
XX N-PSDB; AAY40801.
XX
XX New nucleic acid molecule encoding plant raffinose synthetase -
XX capable of producing raffinose, used as food additives with
XX beneficial effects on gastrointestinal flora
XX
XX Claim 1; Page 31-34; 44pp; English.
XX
XX This sequence represents the soybean raffinose synthetase of the
XX invention. The raffinose synthetase is capable of producing raffinose by
XX combining a D-galactosyl group through an alpha (1-6) bond with a
XX hydroxyl group attached to the carbon atom at position 6 of a D-glucose
XX residue in a sucrose molecule. The DNA can be used to modify metabolism
XX of a host organism by introducing into the host organism or its cell so
XX that the content of the raffinose family oligosaccharides in the host
XX organism or cell is changed. Raffinose oligosaccharides are useful as
XX food additives with beneficial effects on the gastrointestinal flora.
XX
XX Sequence 781 AA;

SQ
Query Match 88.1%; Score 497; DB 19; Length 781;
Best Local Similarity 85.0%; Pred. No. 1.4e-54;
Matches 85; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHFAASRAISGPIYVSDSGKHNFDLKRLVDPDGSILRSEYALP 60
DB 511 PDMDMFQSTHPCAFHFAASRAISGPIYVSDSGKHNFKLKSLALPDGTTILRCQHVALP 570
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGCGWC 610

RESULT 12
AAY30143
ID AAY30143 standard; Protein; 781 AA.
XX
XX AAY30143;
XX
XX 26-OCT-1999 (first entry)
XX
XX Amino acid sequence of a raffinose synthase protein.
XX
XX Raffinose synthase; plant; sucrose; raffinose.
XX
XX Glycine max.
XX
XX JP11215984-A.
XX
XX 10-AUG-1999.
XX
XX 12-DEC-1997; 97JP-0342899.
XX
XX 28-NOV-1997; 97JP-0329006.
XX
XX 18-DEC-1996; 96JP-0338673.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX MPI: 1999-511112/43.
XX
XX N-PSDB; AAZ10002.
XX
XX New raffinose synthase gene - is prepared from a plant material
XX

```

```

XX
XX Claim 8; Page 25-27; 40pp; Japanese.
XX
XX The present sequence represents a raffinose synthase protein. The
XX sequence is isolated from plant material. The protein forms raffinose
XX by complexing alpha(1 to 6) - D-galactosyl hydroxyl group of the 6C
XX of D-glucose residue in sucrose molecules.
XX
XX Sequence 781 AA;

SQ
Query Match 88.1%; Score 497; DB 20; Length 781;
Best Local Similarity 85.0%; Pred. No. 1.4e-54;
Matches 85; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHFAASRAISGPIYVSDSGKHNFDLKRLVDPDGSILRSEYALP 60
DB 511 PDMDMFQSTHPCAFHFAASRAISGPIYVSDSGKHNFKLKSLALPDGTTILRCQHVALP 570
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGCGWC 610

RESULT 13
AAB49400
ID AAB49400 standard; Protein; 781 AA.
XX
XX AAB49400;
XX
XX 07-MAR-2001 (first entry)
XX
XX Soybean raffinose synthase.
XX
XX Plant promoter; transgenic plant; desired property.
XX
XX Glycine max.
XX
XX EP1048733-A2.
XX
XX 02-NOV-2000.
XX
XX 27-APR-2000; 2000EP-0108962.
XX
XX 30-APR-1999; 99JP-0124527.
XX
XX 01-SEP-1999; 99JP-0247211.
XX
XX (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX Ishige F, Watanabe E, Oeda K;
XX
XX MPI: 2001-104537/12.
XX
XX N-PSDB; AAC89523.
XX
XX New soybean plant promoters useful for generating transgenic plants
XX with desired properties.
XX
XX Example 6; Page 24-27; 36pp; English.
XX
XX The present invention provides novel plant promoters which can be used in
XX the production of transgenic plants which express genes with desired
XX properties.
XX
XX Sequence 781 AA;

SQ
Query Match 88.1%; Score 497; DB 22; Length 781;
Best Local Similarity 85.0%; Pred. No. 1.4e-54;
Matches 85; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 1 PDMDMFQSTHPCAFHFAASRAISGPIYVSDSGKHNFDLKRLVDPDGSILRSEYALP 60
DB 511 PDMDMFQSTHPCAFHFAASRAISGPIYVSDSGKHNFKLKSLALPDGTTILRCQHVALP 570
DB 571 TRDCLFEDPLHNGETMLKIWNLNKFTGVIGAFNCGGCGWC 610

```

Db 571 TRDCLFEDPLHDGKTMKIKIMNLKNTYGVGLFNCGGGMC 610

RESULT 14

AA70978
ID AA70978 standard; Protein: 758 AA.

AA70978;

09-AUG-2000 (first entry)

Soybean raffinose synthase from clone sfl1.pk125.d4.

Soybean; raffinose synthase; raffinose saccharide;

clone sfl1.pk125.d4; nutritional; soy protein.

Glycine max.

MO200024915-A2.

04-MAY-2000.

22-OCT-1999; 99MO-US24923.

23-OCT-1998; 98US-0105451.

(DUPO) DU PONT DE NEMOURS & CO E. I.

Allen SM, Hiltz WD;

WPI; 2000-350754/30.

N-PSDB; AAD00335.

Nucleic acids and encoded proteins involved in the biosynthesis of raffinose, useful for producing soybean seeds with a reduced raffinose content and therefore improved nutritional quality -

Claim 2; Page 47-49; 58pp; English.

The present sequence is a raffinose synthase from clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA library sfl1. Raffinose synthase is involved in the biosynthesis of raffinose and higher homologues in the raffinose saccharide family from sucrose. The present sequence is useful for reducing the raffinose saccharide content of soybean seeds which improves the nutritional quality of the soy protein products derived from them.

Sequence 758 AA;

Query Match 87.6%; Score 494; DB 21; Length 758;

Best Local Similarity 84.2%; Pred. No. 3.2e-54;

Matches 85; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

1 PDMDFOGTHPCAAEHAASRAISGPIYVSDVGKHNFDLKKLVLPDGSILRSEYALP 60

487 PDMDFOGTHPCAAEHAASRAISGPIYISDTVGNHNFELKLTALPDGSILRCEHYALP 546

61 TRDCLFEDPLHDGKTMKIKIMNLKNTYGVIGAFNCGGGMCR 101

547 TRDCLFADPLHDGKTMKIKIMNLKNTYGVIGAFNCGGGMCR 587

RESULT 15

AA732073
ID AA732073 standard; Protein: 783 AA.

AA732073;

17-JAN-2000 (first entry)

Sugarbeet raffinose synthase.

KW Raffinose synthase; sugarbeet; transgenic plant.

OS Beta vulgaris.

Key Location/Qualifiers

Misc-difference 227 /note= "encoded by CCW"

Misc-difference 348 /note= "encoded by CCR"

EP953643-A2.

03-NOV-1999.

27-APR-1999; 99EP-0107430.

30-APR-1998; 98JP-0120550.

30-APR-1998; 98JP-0120551.

04-DEC-1998; 98JP-0345590.

10-DEC-1998; 98JP-0351246.

(SUMO) SUMITOMO CHEM CO LTD.

Watanabe E, Oeda K;

WPI; 1999-593144/51.

N-PSDB; AA220208.

New sense and antisense genes, useful for altering the level of raffinose in food plants -

Claim 25; Page 22-24; 55pp; English.

This sequence represents sugarbeet raffinose synthase, a protein that can bind a D-galactosyl group through an alpha(1-6) bond to the hydroxy group attached to the carbon atom at the 6-position of the D-glucose residue in a sucrose molecule to form raffinose. cDNA (see AA220208) encoding the enzyme was isolated from sugarbeet cv. haming leaf cDNA by PCR. Probes or primers generated from plant other raffinose synthase genes (see AA220207-10) may be used to obtain (claimed). These genes may be used to control the levels of raffinose produced in plants. Antisense genes can be used to knock out existing gene activity, and sense genes to increase the level of gene activity. The resulting transgenic plants may be used as a food source to alter the growing conditions for gut enterobacteria, providing general health advantages.

Sequence 783 AA;

Query Match 84.4%; Score 476; DB 20; Length 783;

Best Local Similarity 83.2%; Pred. No. 6.8e-52;

Matches 84; Conservative 7; Mismatches 10; Indels 0; Gaps 0;

1 PDMDFOGTHPCAAEHAASRAISGPIYVSDVGKHNFDLKKLVLPDGSILRSEYALP 60

510 PDMDFOGTHPCAAEHAASRAISGPIYSDVGKHNIPDLKRLVLAOSILRCEHYALP 569

61 TRDCLFEDPLHDGKTMKIKIMNLKNTYGVIGAFNCGGGMCR 101

570 TRDCLFVDPDLHDGKTMKIKIMNLKNTYGVIGAFNCGGGMCR 610

Search completed: March 27, 2003, 10:03:45
Job time : 72.439 secs

RESULT 2
US-09-847-172-23
; Sequence 23, Application US/09847172
; Publication No. US20030007978A1


```
; PRIOR FILING DATE: 2001-01-12
; NUMBER OF SEQ ID NOS: 561
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 266
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Shigella flexneri
US-10-043-487-266

Query Match
Best Local Similarity 26.5%; Pred. No. 27;
Matches 26; Conservative 13; Mismatches 27; Indels 32; Gaps 5;

QY 1 PWDMEFQSTHPCAAFHASRAISGPIYSDSV-----GKNHFDLTKL----- 44
DB 691 PFDLNQNLH-----GSRLVQDVVEYLESTYRPFVSLQDDYFATLRELEATLRTQSL 743
;
QY 45 ---VLPDGSILRSEYALPTRDCLFEDPLHNGETMLKI 79
DB 744 SLAVIPEGKIMNNYY---QECLEFY--LHNYSTNLAI 775
;

RESULT 6
US-09-801-368-148
; Sequence 148, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Call, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 148
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-148

Query Match
Best Local Similarity 11.4%; Score 64.5; DB 10; Length 358;
Matches 24; Conservative 10; Mismatches 24; Indels 33; Gaps 4;

QY 15 FHASRAISGPIYSDVSGKHNFDLTKLVLPDGSIL-----LRSEYAL-----PTRD 63
DB 186 FNA-----YOTANFLQENFEAIKKVYVCPDSCLOQORQPKRPRKRSLLLLIPNAS 235
;
QY 64 CLFEDPLHNGETMLKIWNKFTGVGATGNC 94
DB 236 ELSETP-----LMRFAGVGFPMNC 254
;

RESULT 7
US-10-047-539-2
; Sequence 2, Application US/10047539
; Patent No. US20020177547A1
```

```
; GENERAL INFORMATION:
; APPLICANT: MOLLING, KARIN
; APPLICANT: PAVLOVIC, JOVAN
; APPLICANT: NAMRATH, MICHAEL
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS FOR TREATING OR PREVENTING
; FILE REFERENCE: VOS-27
; CURRENT APPLICATION NUMBER: US/10/047,539
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: EP 01 10 0914.9
; PRIOR FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO: 2
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-047-539-2

Query Match
Best Local Similarity 11.4%; Score 64.5; DB 9; Length 626;
Matches 29; Conservative 17; Mismatches 42; Indels 29; Gaps 7;

QY 1 PWDMEFQSTHPCAAFHASRAISGPIYV-----SDVSGKHNFDLTKLVLPDGSIL----- 52
DB 50 PWTVEVQSGNCWRGQVSLRVINDGPTLVGANASFIALH--PPGSOK-VLPDGOVIMANN 107
;
QY 53 -----RSEYALPTRDCLFED--PLHNG-----ETMLKIWN--NKFTGVIG 90
DB 108 TIINSQVWGQPYVPQEPDADCVPPDGGCPSPRPRPRSRFVYWKTKGYOVLG 164
;

RESULT 8
US-10-081-281-93
; Sequence 93, Application US/10081281
; Patent No. US20020151707A1
; GENERAL INFORMATION:
; APPLICANT: Kindsvogel, Wayne
; APPLICANT: Gross, Jane A.
; APPLICANT: Sheppard, Paul
; TITLE OF INVENTION: Immune Mediators and Related Methods
; NUMBER OF SEQUENCES: 121
; CORRESPONDENCE ADDRESS:
; ADDRESS: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/081,281
; FILING DATE: 20-Feb-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/261,811A
; FILING DATE: 03-Mar-1999
; APPLICATION NUMBER: US 08/480,002
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/482,133
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 08/483,241
; FILING DATE: 07-JUN-1995
; APPLICATION NUMBER: US 60/005,964
; FILING DATE: 27-OCT-1995
; APPLICATION NUMBER: US 08/657,581
; FILING DATE: 07-JUN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Parent, Annette S.
```

```

      REGISTRATION NUMBER: 42.,058  

      REFERENCE/DOCKET NUMBER: 014058-005630US  

TELECOMMUNICATION INFORMATION:  

      TELEPHONE: (415) 576-0200  

      TELEFAX: (415) 576-0300  

INFORMATION FOR SEQ ID NO: 93:  

    SEQUENCE CHARACTERISTICS:  

        LENGTH: 207 amino acids  

        TYPE: amino acid  

        TOPOLOGY: linear  

MOLECULE TYPE: protein  

SEQUENCE DESCRIPTION: SEQ ID NO: 93:  

US-10-081-281-93  
  

Query Match           11.3%; Score 63.5; DB 12; Length 207;  

Best Local Similarity   22.4%; Pred.No.8.8;  

Matches 19; Conservative 18; Mismatches 25; Indels 23; Gaps 4,  
  

QY       31 DSVGHHNFDLKKVLVLPDGS-----ILRSEYYALPTR-----DCLEDPLH 71  

          |:|::: : |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  

Db       104 DTYCHNHNGVVESTFWOQSASAGIKEEHVIIOAEFYLVLPDQSGEFMFDFDGDEIFHVDMA 163  
  

OY       72 NGETMLKIMNLKFTGVIGAFMCGC 96  

          ||:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|  

Db       164 KKET---VWRLEEFGRAFSFAQG 184
```

RESULT 9
 US-09-843-676-55
 : Sequence 55, Application US/09843676
 : Patent No. US20020164786A1
 :
 : GENERAL INFORMATION:
 :
 : APPLICANT: Cech, Thomas R.
 : Lingner, Joachim
 : Nakamura, Toru
 : Chapman, Karen B.
 : Morin, Gregg B.
 : Harley, Calvin
 : Andrews, William H.
 :
 : TITLE OF INVENTION: NO. US20020164786A1el Telomerase
 :
 : NUMBER OF SEQUENCES: 225
 :
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Townsend and Townsend and Crew LLP
 : STREET: Two Embarcadero Center, 8th Floor
 : CITY: San Francisco
 : STATE: California
 : COUNTRY: United States of America
 : ZIP: 94111
 :
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.30
 :
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/843,676
 : FILING DATE: 26-Apr-2001
 : CLASSIFICATION: 536
 :
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: US/08/854,050
 : FILING DATE: 09-May-1997
 : APPLICATION NUMBER: US 08/846,017
 : FILING DATE: 25-Apr-1997
 : APPLICATION NUMBER: US 08/844,419
 : FILING DATE: 18-Apr-1997
 : APPLICATION NUMBER: US 08/724,643
 : FILING DATE: 01-Oct-1996
 :
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Apple, Randolph T.
 : REGISTRATION NUMBER: 36,429
 : REFERENCE/DOCKET NUMBER: 015389-002930US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (415) 576-0200
 : TELEFAX: (415) 576-0300

```

? INFORMATION FOR SEQ ID NO: 55:
?     SEQUENCE CHARACTERISTICS:
?         LENGTH: 884 amino acids
?         type: amino acid
?     STRANDEDNESS: No. US20020164786A1 Relevantant
?     TOPOLOGY: No. US20020164786A1 Relevantant
?     MOLECULE TYPE: peptide
?     SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-843-676-55

```

Query Match	11.2%	Score 63;	DB 9;	Length 884;
Best Local Similarity	29.5%	Pred. No. 59;		
Matches	23;	Conservative	15;	Mismatches 30;
				Indels 10;
				Gaps 4;
Qy	26	PIYVSDSGKHNFDLKKI--VLDPGSLI----	RSEYALPRDC--LEEDPL--HNGET	75
Db	499	PQLQIDRLKEPQRRLKKFNNTLP	PELYEMKPFVYKCYDIP	PMECMRILKDAKKNENGEF
				558
Qy	76	MLKTNLKKFTGYIGAE	N 93	
Db	559	VRSQIEFNITGVLEKLE	N 576	

```

US-097-766-253-55
: Sequence 55, Application US/09766253
: Publication No. US20020187471A1
: GENERAL INFORMATION:
: APPLICANT: Cech, Thomas R.
:           Lingner, Joachim
:           Nakamura, Toru
:           Chapman, Karen B.
:           Morin, Gregg B.
:           Harley, Calvin
:           Andrews, William H.
: TITLE OF INVENTION: No. US20020187471A1e1 Telomerase
: NUMBER OF SEQUENCES: 171
: CORRESPONDENCE ADDRESS:
: ADDRESS: Townsend and Townsend and Crew LLP
: STREET: Two Embarcadero Center, 8th Floor
: CITY: San Francisco
: STATE: California
: COUNTRY: United States of America
: ZIP: 94111
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/766,253
: FILING DATE: 19-Jan-2001
: CLASSIFICATION: <unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/846,017
: FILING DATE: 1997-04-25
: APPLICATION NUMBER: US 08/724,643
: FILING DATE: 01-OCT-1996
: ATTORNEY/AGENT INFORMATION:
: NAME: Apple, Randolph T.
: REGISTRATION NUMBER: 36,429
: REFERENCE/DOCKET NUMBER: 015389-00292005
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 55:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 884 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: peptide
: SEQUENCE DESCRIPTION: SEQ ID NO: 55:

```

US-09-766-253-55

Query Match 11.2% Score 63; DB 9; Length 884;

Best Local Similarity 29.5% Pred. No. 59;

Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4;

OY 26 PIVSDSVCKHNFDLKKL-VLPDGSIL- -RSEYVALPTRDC--LFEDPL--HNGET 75

DB 499 PFOIADRIKEFKORLLKKNVLPPELYFMKFDVKSCTSDSPRMECKRIILKDALKNNGFF 558

OY 76 MLKIWNLNKFTGVIGAFN 93

DB 559 VRSQYFNTNTGVKLKFN 576

RESULT 11

US-09-438-486-55

Sequence 55, Application US/09438486

Publication No. US2003009019A1

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

APPLICANT: Lingner, Joachim

APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.

APPLICANT: Morlin, Gregg B.

APPLICANT: Harley, Calvin

APPLICANT: Andrews, William H.

TITLE OF INVENTION: No. US2003009019A1el Telomerase

NUMBER OF SEQUENCES: 223

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/438,486

FILING DATE: 12-NOV-1999

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/851,843

FILING DATE: 06-MAY-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002931US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 884 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: peptide

US-09-438-486-55

Query Match 11.2% Score 63; DB 9; Length 884;

Best Local Similarity 29.5% Pred. No. 59;

Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4;

OY 26 PIVSDSVCKHNFDLKKL-VLPDGSIL- -RSEYVALPTRDC--LFEDPL--HNGET 75

DB 499 PFOIADRIKEFKORLLKKNVLPPELYFMKFDVKSCTSDSPRMECKRIILKDALKNNGFF 558

OY 76 MLKIWNLNKFTGVIGAFN 93

DB 559 VRSQYFNTNTGVKLKFN 576

RESULT 12

US-10-053-758-55

Sequence 55, Application US/10053758

Publication No. US20030032075A1

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

APPLICANT: Lingner, Joachim

APPLICANT: Nakamura, Toru

APPLICANT: Chapman, Karen B.

APPLICANT: Morlin, Gregg B.

APPLICANT: Harley, Calvin

APPLICANT: Andrews, William H.

TITLE OF INVENTION: No. US20030032075A1el Telomerase

NUMBER OF SEQUENCES: 225

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/053,758

FILING DATE: 18-Jan-2002

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/854,050

FILING DATE: 09-MAY-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/851,843

FILING DATE: 06-MAY-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 884 amino acids

TYPE: amino acid

STRANDEDNESS: No. US20030032075A1 Relevant

TOPOLOGY: No. US20030032075A1 Relevant

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-10-053-758-55

Query Match 11.2%; Score 63; DB 9; Length 884;

Best Local Similarity 29.5%; Pred. No. 59;

Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4;

QY 26 PIVSDSVGKHNFDLKKL--VLPDGSIL---RSEYYALPTRDC--LFEDPL--HNGET 75

DB 499 PTQIADRIKEFKORLLKFNNVLPETYFMKFDVKSCTDIPRMECMRLKDALKNENGFF 558

QY 76 MLKIMNLKFTGYGAFN 93

DB 559 VRSQYFNTMTGVKLNF 576

RESULT 13

US-10-054-295-55

Sequence 55, Application US/10054295

Publication No. US20030044953A1

GENERAL INFORMATION:

APPLICANT: Cech, Thomas R.

Lingner, Joachim

Nakamura, Toru

Chapman, Karen B.

Morin, Gregg B.

Harley, Calvin

Andrews, William H.

TITLE OF INVENTION: No. US20030044953A1el Telomerase

NUMBER OF SEQUENCES: 225

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP

STREET: Two Embarcadero Center, 8th Floor

CITY: San Francisco

STATE: California

COUNTRY: United States of America

ZIP: 94111

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/054,295

FILING DATE: 18-Jan-2002

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/854,050

FILING DATE: <Unknown>

APPLICATION NUMBER: US 08/846,017

FILING DATE: 25-APR-1997

APPLICATION NUMBER: US 08/844,419

FILING DATE: 18-APR-1997

APPLICATION NUMBER: US 08/724,643

FILING DATE: 01-OCT-1996

ATTORNEY/AGENT INFORMATION:

NAME: Apple, Randolph T.

REGISTRATION NUMBER: 36,429

REFERENCE/DOCKET NUMBER: 015389-002930US

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 576-0200

TELEFAX: (415) 576-0300

INFORMATION FOR SEQ ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 884 amino acids

TYPE: amino acid

STRANDEDNESS: No. US20030044953A1 Relevant

TOPOLOGY: No. US20030044953A1 Relevant

MOLECULE TYPE: Peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 55:

US-10-054-295-55

Query Match 11.2%; Score 63; DB 9; Length 884;
Best Local Similarity 29.5%; Pred. No. 59;

Matches 23; Conservative 15; Mismatches 30; Indels 10; Gaps 4;

QY 26 PIVSDSVGKHNFDLKKL--VLPDGSIL---RSEYYALPTRDC--LFEDPL--HNGET 75

DB 499 PTQIADRIKEFKORLLKFNNVLPETYFMKFDVKSCTDIPRMECMRLKDALKNENGFF 558

QY 76 MLKIMNLKFTGYGAFN 93

DB 559 VRSQYFNTMTGVKLNF 576

RESULT 14

US-09-847-172-44

Sequence 44, Application US/09847172

Publication No. US20030007978A1

GENERAL INFORMATION:

APPLICANT: OREGON HEALTH AND SCIENCES UNIVERSITY

APPLICANT: BURROWS, GREGORY G.

APPLICANT: VANDENBARK, ARTHUR A.

TITLE OF INVENTION: RECOMBINANT MHC MOLECULES USEFUL FOR MANIPULATION OF ANTIGEN-S

FILE REFERENCE: 899-58137

CURRENT APPLICATION NUMBER: US/09/847,172

CURRENT FILING DATE: 2001-05-01

PRIOR APPLICATION NUMBER: US 60/200,942

PRIOR FILING DATE: 2000-05-01

PRIOR APPLICATION NUMBER: US 09/153,586

PRIOR FILING DATE: 1998-09-15

PRIOR APPLICATION NUMBER: US 60/064,555

PRIOR FILING DATE: 1997-10-10

PRIOR APPLICATION NUMBER: US 60/064,552

PRIOR FILING DATE: 1997-09-16

NUMBER OF SEQ ID NOS: 44

SOFTWARE: Patentin version 3.1

SEQ ID NO 44

LENGTH: 210

TYPE: PRT

ORGANISM: Homo sapiens

US-09-847-172-44

Query Match 11.0%; Score 62; DB 9; Length 210;

Best Local Similarity 23.2%; Pred. No. 13;

Matches 19; Conservative 18; Mismatches 25; Indels 20; Gaps 4;

QY 31 DSVGKHN-----FDLKLKLVLPDGSILRSEYYALPTR-----DCLFEDPLHNGE 74

DB 107 DTGCRHNYGVVESTYQRRVYKEHVLIQAFLNPQSGEFMFDDGDEIFVYDMAKKE 166

QY 75 TMLKIMNLKFTGYGAFNCG 96

DB 167 T--VWRLEEF-GRFASFEAG 184

RESULT 15

US-09-960-226-4

Sequence 4, Application US/09960226

Patent No. US20020172670A1

GENERAL INFORMATION:

APPLICANT: Rose, David

APPLICANT: Kuntz, Douglas

APPLICANT: Van Den Elsen, Jean

TITLE OF INVENTION: MANNOSEDASE STRUCTURES

FILE REFERENCE: 12243.19US01

CURRENT APPLICATION NUMBER: US/09/960,226

CURRENT FILING DATE: 2001-09-21

PRIOR APPLICATION NUMBER: US 60/263,458

PRIOR FILING DATE: 2001-01-23

PRIOR APPLICATION NUMBER: US 60/234,879

PRIOR FILING DATE: 2000-09-22

NUMBER OF SEQ ID NOS: 9

SOFTWARE: Patentin version 3.1

SEQ ID NO 4

LENGTH: 1096

Fri Apr 4 12:31:34 2003

us-09-675-208-5_copy_510_610.rabb

Page 7

```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-960-226-4

```

Query Match	10.7%	Score 60.5;	DB 9;	Length 1096;
Best Local Similarity	27.6%	Pred. No. 1.5e+02;		
Matches	21;	Conservative	11;	Mismatches 21;
			Indels	23;
			Gaps	4;

```

Oy      41 LKTLVPGG-----SLRSEYALPRTDCL--EDPLHNGET--MLKIMWLN 83
      | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      29 LSESVEDGGPKRSQSNSFGQAGSHLLPQLSSVDYADCLFASQSSGSHNSDVMLEYSLI 88
      | : | : | : | : | : | : | : | : | : | : | : | : | : |
Oy      84 KFTGYICAFNCGGCGW 99
      | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      89 SFD-----NPDGQW 98
      | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

Search completed: March 27, 2003, 10:20:38
Job time : 24.4024 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:07:51 ; Search time 41 Seconds
(without alignments)
3940.023 Million cell updates/sec

Title: US-09-675-208-5

Perfect score: 4225

Sequence: 1 MAPSFKNGSNVVSFDGLND.....QVPWPDSSGSGISVIEYLF 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp-archaea:*
- 2: sp-bacteria:*
- 3: sp-fungi:*
- 4: sp-human:*
- 5: sp-invertebrate:*
- 6: sp-mammal:*
- 7: sp-mnc:*
- 8: sp-organella:*
- 9: sp-phage:*
- 10: sp-plant:*
- 11: sp-rodent:*
- 12: sp-virus:*
- 13: sp-vertebrate:*
- 14: sp-unclassified:*
- 15: sp-virus:*
- 16: sp-bacteriap:*
- 17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4325	100.0	784	10 Q92T62	Q92T62 cucumis sat
2	2945	69.7	783	10 Q9FND9	Q9FND9 arabidopsis
3	2807.5	66.4	798	10 Q8VWV6	Q8VWV6 pisum sativ
4	2643	62.6	816	10 Q942N6	Q942N6 oryza sativ
5	1914.5	45.3	853	10 Q93XK2	Q93XK2 pisum sativ
6	1914	45.3	857	10 Q9SBZ0	Q9SBZ0 phaseolus a
7	1868.5	44.2	863	10 Q8RW08	Q8RW08 strachys aff
8	1853.5	43.9	807	10 Q9SVY4	Q9SVY4 arabidopsis
9	1482	35.1	757	10 Q40077	Q40077 hordeum vul
10	1473.5	34.9	765	10 Q43408	Q43408 brassica o1
11	1459.5	34.5	773	10 Q94A08	Q94A08 arabidopsis
12	1458.5	34.2	773	10 Q9SCM1	Q9SCM1 arabidopsis
13	1446.5	34.2	844	10 Q8RX87	Q8RX87 persica amer
14	1413.5	33.5	779	10 Q9M4M7	Q9M4M7 arabidopsis
15	1140	27.0	1170	10 Q9LFE7	Q9LFE7 arabidopsis
16	697.5	16.5	386	10 Q9M442	Q9M442 cicer ariet

17	514.5	12.2	283	10 Q9FWM2	Q9FWM2 arabidopsis
18	490.5	11.6	357	10 Q94666	Q94666 cicer ariet
19	471.5	11.2	371	10 Q04607	Q04607 arabidopsis
20	408.5	9.7	702	2 Q93CM6	Q93CM6 bifidobacte
21	397	9.4	674	17 Q96X62	Q96X62 sulfolobus
22	365	8.6	649	17 Q97094	Q97094 sulfolobus
23	350	8.3	125	10 Q42099	Q42099 arabidopsis
24	219	5.2	104	10 Q8W2G5	Q8W2G5 oryza sativ
25	161	3.8	204	10 Q9AV63	Q9AV63 oryza sativ
26	145	3.4	88	10 Q42092	Q42092 arabidopsis
27	121.5	2.9	538	16 P74506	P74506 synechocyst
28	114.5	2.7	776	2 Q8RM03	Q8RM03 xanthobacte
29	114	2.7	623	5 Q61391	Q61391 caenorhabdi
30	114	2.7	623	5 Q95ZJ2	Q95ZJ2 caenorhabdi
31	113.5	2.7	632	3 Q04049	Q04049 saccharomyc
32	112.5	2.7	1963	10 Q94HY2	Q94HY2 oryza sativ
33	111	2.6	741	10 Q9LNO1	Q9LNO1 arabidopsis
34	106	2.5	4074	4 Q8TC29	Q8TC29 homo sapien
35	105.5	2.5	473	16 Q8XM38	Q8XM38 clostridium
36	105.5	2.5	568	3 Q9P8X6	Q9P8X6 aspergillus
37	104.5	2.5	736	16 Q92M14	Q92M14 rhizobium m
38	104	2.5	482	10 Q65328	Q65328 nicotiana g
39	104	2.5	689	17 Q976H7	Q976H7 sulfolobus
40	104	2.5	933	10 Q8WSR2	Q8WSR2 arabidopsis
41	103.5	2.4	1896	12 Q04232	Q04232 mucosal dis
42	103	2.4	3944	5 Q18667	Q18667 caenorhabdi
43	102.5	2.4	544	16 Q98RA7	Q98RA7 mycoplasma
44	102.5	2.4	582	16 P73732	P73732 synechocyst
45	102.5	2.4	1093	10 Q9LFE6	Q9LFE6 arabidopsis

ALIGNMENTS

RESULT 1

ID Q92T62 PRELIMINARY: PRT: 784 AA.

AC Q92T62;

DT 01-MAY-1999 (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)

DE Rafinose synthase (EC 2.4.1.82).

DE RFS.

GN Cucumis sativus (Cucurbit).

OS Cucumis sativus (Cucurbit).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC Cucurbitales; Cucurbitaceae; Cucumis.

OX NCBI_TaxID=3659;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=LEAF;

RA Ohsumi C., Nozaki J., Kida T.;

RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF073744; AAD02832.1; -

KW Glycosyltransferase; Transferase.

SQ SEQUENCE 784 AA: 86920 MW: 38064491F0908933 CRC64;

Query Match 100.0%; Score 4225; DB 10; Length 784;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSFKNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVNPENIVASPPYISDKS 60

DB 1 MAPSFKNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVNPENIVASPPYISDKS 60

QY 61 PVSVCGEVGFDASEPDSSRHVVSIGKLKDIRFMSIFRFKVMWTTHWVGRNGDLESTQIV 120

DB 61 PVSVCGEVGFDASEPDSSRHVVSIGKLKDIRFMSIFRFKVMWTTHWVGRNGDLESTQIV 120

QY 121 ILEKSDSGRPVFLPIVGGPRTSTQPGDDPFVQVCSGSSKVVDAFBSMLYLHAGD 180

DB 121 ILEKSDSGRPVFLPIVGGPRTSTQPGDDPFVQVCSGSSKVVDAFBSMLYLHAGD 180

```
QY 181 DPAALVAKAMKIVRTHLGTFRLLBECTPPECIVDKFGKCTMDAFYLVHPOGVIEGRHLY 240
    |||
Db 181 DPAALVAKAMKIVRTHLGTFRLLBECTPPECIVDKFGKCTMDAFYLVHPOGVIEGRHLY 240
QY 241 DGGCGPELVILIDGMOSIGHSDPITKEGMNOMYAGBOMCRLKPEANKKFRDYVAPKA 300
    |||
Db 241 DGGCGPELVILIDGMOSIGHSDPITKEGMNOMYAGBOMCRLKPEANKKFRDYVAPKA 300
QY 301 TGPBAGGKMKAFIDELKGEFTEVHHYVMAHLCGYGGLRPQVGPPEARVIOPLVSPG 360
    |||
Db 301 TGPBAGGKMKAFIDELKGEFTEVHHYVMAHLCGYGGLRPQVGPPEARVIOPLVSPG 360
QY 361 LQMTMEDLAVDKIVLHKVGLVPEPKAEEMTEGLHAHLEKVIDGVKIDVTHLLEMCEDY 420
    |||
Db 361 LQMTMEDLAVDKIVLHKVGLVPEPKAEEMTEGLHAHLEKVIDGVKIDVTHLLEMCEDY 420
QY 421 GGRVDLAKAYKAMTKSINKHFGKNGVYASMEHCNDPMFLGTEAISLGRVGDDPWCCTDPS 480
    |||
Db 421 GGRVDLAKAYKAMTKSINKHFGKNGVYASMEHCNDPMFLGTEAISLGRVGDDPWCCTDPS 480
QY 481 GDBNGTFMLOGCHVHCANDSLMWGNFIHPDWMFOSTHPCAAFHAASRAISGGPIYVD 540
    |||
Db 481 GDBNGTFMLOGCHVHCANDSLMWGNFIHPDWMFOSTHPCAAFHAASRAISGGPIYVD 540
QY 541 SVKRNHNDLKLKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETMLKINLNKFTGTGTA 600
    |||
Db 541 SVKRNHNDLKLKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETMLKINLNKFTGTGTA 600
QY 601 FNCGGGMCRETRNOCFSQYSKRVTSKTPKDIEMHSGENPISIEGVKFALYLYOAKK 660
    |||
Db 601 FNCGGGMCRETRNOCFSQYSKRVTSKTPKDIEMHSGENPISIEGVKFALYLYOAKK 660
QY 661 LILSKPSQDIDLADPEFELITVSPVTKLIQTSLHAPIGLVNMLNTSGAISVDYDD 720
    |||
Db 661 LILSKPSQDIDLADPEFELITVSPVTKLIQTSLHAPIGLVNMLNTSGAISVDYDD 720
QY 721 LSSVEIGVKGCGEMRVASAKPRACRIDGEGVGFXYDDQDQVVVQVWPIIDSSGGSIVT 780
    |||
Db 721 LSSVEIGVKGCGEMRVASAKPRACRIDGEGVGFXYDDQDQVVVQVWPIIDSSGGSIVT 780
QY 781 EYLF 784
    |||
Db 781 EYLF 784
    |||

RESULT 2
Q9FND9 PRELIMINARY: PRT: 783 AA.
AC Q9FND9:
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Raffinose synthase protein.
GN MPO12.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT *Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.*;
RL DNA Res. 4:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
```

```
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shin P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shin P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB006702; BAB1595.1; -
DR EMBL: AY062781; AAL32859.1; -
DR EMBL: AY081645; AAM10207.1; -
SQ SEQUENCE 783 AA: 86237 MW: 3C37D1D7871888AC CRC64;

Query Match 69.7%; Score 2945; DB 10; Length 783;
Best Local Similarity 69.2%; Pred. No. 4.7e-231;
Matches 544; Conservative 99; Mismatches 117; Indels 26; Gaps 9;

QY 14 SPFGN--DMSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS---PVSGCF 67
    |||
Db 9 SDGGINGVDFTEKFRLEDSTLLANGOVYLDVNVNLTSSPVL-VKDGVPVLDVSGSF 67
QY 68 VGFDA-SEPSRHVVSIGKLDIRFMSIFRFKYAMTTTHWGRNGDLESEFQVILEK-- 124
    |||
Db 68 IGFNLDEPSSHVVASIGKLNIRFMSIFRFKYAMTTTHWGRNGDLESEFQVILEK-- 127
QY 125 -----SDSGRPVFLPLIVEGPRFISIQPDDEFDVDCYESSGSKVYDASFSMLYLA 178
    |||
Db 128 SDSGRPGSGSRPVVLLPLLEGFSRFSFGEDDDVAVCSGSETEVTSGEFQIYVHA 187
QY 179 GDDPFALVAKAMKIVRTHLGTFRLLBECTPPECIVDKFGKCTMDAFYLVHPOGVIEGRH 238
    |||
Db 188 GDDPFALVAKAMKIVRTHLGTFRLLBECTPPECIVDKFGKCTMDAFYLVHPOGVIEGRH 247
QY 239 LVDGCGPELVILIDGMOSIGHSDPITKEGMNOMYAGBOMCRLKPEANKKFRDYVAPKA 298
    |||
Db 248 LVDGCGPELVILIDGMOSIGHSDPITKEGMNOMYAGBOMCRLKPEANKKFRDYVAPKA 307
QY 299 KATGPBAGGKMKAFIDELKGEFTEVHHYVMAHLCGYGGLRPQVGPPEARVIOPLVSPG 358
    |||
Db 308 K----DQNDVMKAFVRLDDEFSITVDIYVMAHLCGYGGLRPQVGPPEARVIOPLVSPG 363
QY 359 PGLQMTMEDLAVDKIVLHKVGLVPEPKAEEMTEGLHAHLEKVIDGVKIDVTHLLEMCEDY 418
    |||
Db 364 PGLQMTMEDLAVDKIVLHKVGLVPEPKAEEMTEGLHAHLEKVIDGVKIDVTHLLEMCEDY 423
QY 419 DYSGRYDLAKAYKAMTKSINKHFGKNGVYASMEHCNDPMFLGTEAISLGRVGDDPWCCTD 478
    |||
Db 424 KYGGRVDLAKAYKAMTKSINKHFGKNGVYASMEHCNDPMFLGTEAISLGRVGDDPWCCTD 483
QY 479 PSCDPMCTFWLQGCCHVHCANDSLMWGNFIHPDWMFOSTHPCAAFHAASRAISGGPIY 538
    |||
Db 484 PSCDPMCTFWLQGCCHVHCANDSLMWGNFIHPDWMFOSTHPCAAFHAASRAISGGPIY 543
QY 539 SDSVGKHNPDLLKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETMLKINLNKFTGTGTA 598
    |||
Db 544 SDSVGKHNPDLLKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETMLKINLNKFTGTGTA 603
QY 599 GAFNCGGGGCRETRNOCFSQYSKRVTSKTPKDIEMHSGENPISIEGVKFALYLYOAKK 658
    |||
Db 604 GAFNCGGGGCRETRNOCFSQYSKRVTSKTPKDIEMHSGENPISIEGVKFALYLYOAKK 663
QY 659 KKLILSKPSQDIDLADPEFELITVSPVTKLIQTSLHAPIGLVNMLNTSGAISVDYDD 718
    |||
Db 664 KKLILSKPSQDIDLADPEFELITVSPVTKLIQTSLHAPIGLVNMLNTSGAISVDYDD 723
QY 719 DDLSSVEIGVKGCGEMRVASAKPRACRIDGEGVGFXYDDQDQVVVQVWPIIDSSGGS 778
    |||
```

Db	724	DE--SVEGVFCAGSEFRVYASKKPVSCSIDGEEVVEBFG--EDSMWVWQVPMV---	SSPDDGLS	777
Qy	779	VIEIYLF 784		
Db	778	SIQYLF 783		
RESULT 3				
ID	Q8VWN6	PRELIMINARY:	PRT:	798 AA.
AC	Q8VWN6:			
DT	01-MAR-2002 (Tremblrel. 20, Created)			
DT	01-MAR-2002 (Tremblrel. 20, last sequence update)			
DT	01-MAR-2002 (Tremblrel. 20, last annotation update)			
DE	Raffinose synthase (EC 2.4.1.82).			
GN	Pisum			
OS	Pisum sativum (Garden pea).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosid 1; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.			
OX	NCBI_TextID=3688;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. WUNDER VON KEIVEDON; TISSUE=SEED;			
RA	Peterbauer T., Mach L., Mucha J., Richter A.;			
RT	"Molecular characterization of raffinose synthase from pea (Pisum			
RT	sativum L.) seeds."			
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AJ426475; CAD20127.2; -			
KW	Transferase; Glycosyltransferase.			
SO	SEQUENCE 798 AA; 88717 MW; 803F3ED5BF8617B7 CRC64;			
Query Match	66.4%; Score 2807.5; DB 10; Length 798;			
Best Local Similarity	65.1%; Pred. No. 7.8e-220;			
Matches 510; Conservative 120; Mismatches 125; Indels 29; Gaps 10;				
Qy	22	SSP---FAIDGS-DFTVNGHSLSDVPEMI-----VASPSPTYSIDKSPVSV-----	64	
Db	23	NSPLSLISIDDSGRNLVGVNHPFLTVGPPIITTTTSTPSLPDFKSNKTTJANNNTLQO	82	
Qy	65	-GCFVGFDASEDSRHVYSIGLKIDIRFMSIRFVYMMTHVGNKRGDLESEFOVILE	123	
Db	83	QGCFVGFMTTEKSHHVVPLGKLGKIGKFTSIRFKYMMTHVGNKRGDLEHLOHETOLIID	142	
Qy	124	KSDS-GREYVFLPIVEGPFRTSIOPGDDDFVDCVSESSKSVVASPSRSMYLIAGDDP	182	
Db	143	KNISLGRPYVLLPLILSNSFRSLQPLGUNDYDMSVSESSHTVGTSTFACILYHLNSDP	202	
Qy	183	FALVEAMKIYVTHLCTRLLEKTRPGIIVDKFGCMCTDAFTLVNPGVIEGVRHLVDG	242	
Db	203	YALVEAVKVIOTKLTGTEKTRPSLIEKFGCTWPAFLKLVHPKGVWGEYKALTDG	262	
Qy	243	GRRPGVILIDDCWOSIGH--DSDPIT--KESMNTVAGEOMPCGLLKEOENYKRDVYNPRA	300	
Db	263	GCRPGFVILIDDCWOSISHDDDPVTERDGMNNTSAGEOMPCRLIYEENYKREYEN---	319	
Qy	301	TGPRAGQGMKMAFIDELGGEFTVTHVYVMAHALCGYWGGLRQVPGLPPEARVIOPLSPG	360	
Db	320	-GDNGKKGGLGVFVNDKKEFRSVESVYVMAHALCGYWGVRKVCGMPEAKKVVYKLSFG	378	
Qy	361	LQMTMEDLAVDKIVLHKKGLVPEKAEMEYBSGLAHLEKVGIDGVKIDYIHLLEMLCEDY	420	
Db	379	YKMTMEDLAVDKIVRNGVGLVPRNLAOEEMFDIHSLSAESAGIDGVKVDVYIHLLELLESEY	438	
Qy	421	GRRVLDARAKYKAAFKSIKKHFKGNGVIASMHCNDPMFLGTAEVSLGVGVDPMCTDS	480	
Db	439	GRRVELAKYKAAFTSSVKKHFKGNGVIASMHCNDPFLGTAEVSLGVGVDPMCTDS	498	
Qy	481	GDPNSTFWLQCGAHVHCANDSLMGNFIIPDMDCQSTHPCAAFIHAASRAISGGPIYSD	540	
Db	499	GDPNSTFWLQCGAHVHCAYNSLMMGNFIIPDMDCQSTHPCAEFIHAASRAISGGPIYSD	558	
Qy	541	SVGKNHFDLKLVLPLDGSILNSEYALPTDCLREEDPLHNGETMLKTVNLNKFVGIVA	600	

Dd		559	CAGNNHFKLLKSFVLPDGSILRCQHYALPTRCLEFEDPLHNCKTMYLKNLKAYAGVLG	618
Oy		601	FNCGGGCMCRETRNQCESQYSKYRVTSKTNPKDIEMWSEGNDSIEGVTFYALLYOAKK	660
Dd		619	FNCGGGCMCPETRRKKASASEFSHAVTCVASPEDIEMCNCKTPMDJKGVDFAVYFEKKEK	678
Oy		661	LILSPSODLDIALDPPEELITVSPVKLIOTSLHFAPILGLVMNLNSGAIOADYDDDD	720
Dd		679	LILMCSORLEVESLEPESFELMTVSPLKFSSRLIOFAPIGLVMNLNSGVAOSLEFDOS	738
Oy		721	LSVVEIGVGCCEMVPAASKRPACRIODEGEPFYDDOMVVVOYPMPIDSSSGISLYI	780
Dd		739	ASLVYIGVHGCGELSVFASEKFPVCKIKDSVSEVPDI-EDKMRYVOLIMP--GSSLISLV	794
Oy		781	EYLFF 784	
Dd		795	EFLFF 798	
 RESULT 4 0942t6 PRELIMINARY; PRT: 816 AA.				
ID	0942t6			
AC	0942t6			
Dt	01-DEC-2001 (TREMBLrel. 19, Created)			
Dt	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
Dt	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			
DE	Putative raffinose synthase.			
GN	P0583G08.2.			
OS	Oryza sativa (Rice).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;			
OC	Echaritoidae; Oryzaceae; Oryza.			
OX	NCBITaxId=45330;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV. NIPPONBARE;			
RA	Sasaki T., Matsunoto T., Yamamoto K.;			
RT	"Oryza sativa nipponbare(Ga3) genomic DNA, chromosome 1, PAC			
RT	clone:P0583G08."			
RL	Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.			
DR	EMBL: AP003282; BAB64768.1..			
SO	SEQUENCE 816 AA; 89388 MW; B316EDEF3566C5178 CRC64;			
 Query Match 62.6%; Score 2643; DB 10; Length 816; Best Local Similarity 59.6%; Pred.No.2,le=206; Matches 500; Conservative 111; Mismatches 150; Indels 78; Gaps 16				
Oy	1 MAPSRKNG-----SNVVSFDGLNDMSSP-FAIDGSDFTVNGHSFLSDPVENIVASPSPYT	55		
Dd	1 MAPNLSKARDKDLIGDVAAVDGL-IKPRPFTLGKGLDAVGHPFLDLDPANIRLTGPS-T	57		
Oy	56 SLDKSPV---SVGCPVGPDPASERPDSHNVVISICKLDIRMSIFRFRVMMTTIMVGENGD	112		
Dd	58 LVPNSDVPPAAGSFLGFCDPAKADKHVVRICKLRDTFRMSIFRFVMMTTIMVGTNGRD	117		
Oy	113 LESEOTIVLEKSDS-----GRPYFLPLIVEGPFRTSIQPDP-DDEVDCVESSESRY	165		
Dd	118 VENTQMMLDSDGTSKSSTGPRPYVLLPLRYEGPRACLDESKADYVNMHLESSSYV	177		
Oy	166 VDASERSMLYLHAGDDPALVLEAMKIYTHLGTFRLLBEKTIPRGIVDKFGCTWDAPYL	225		
Dd	178 RGSVFSAUYLAHADDPEDLVKRDAMRVVANHTGFRIMEEKTRPPYVDKFGCTWDAPYL	237		
Oy	226 TVHPGVIQGVNHVNDGCGRPGVLYLIDCWOSTGHUDDSPITK--ECMNQTVAQEOPRCY	283		
Dd	238 KYNPBSVWGVRRLADGGCRPGVLYLIDDWSQICHDDDLGSCAGEBMNTSAGEOQRCL	297		
Oy	284 LKFOENEYKERDYVNPKATPRAGOKGMKAFFIDELKEGEFTVEHVYMHALCGYMGCLRPO	343		
Dd	298 INFOENYKRREY-----KSGMSGFVDEMKAFFTEDEVYVYMHALLCGYMGLRPG	346		
Oy	344 VQGLPEARVYIQVLSFGLOMTMEDLAVDKTVLHKVGLVPRPEKAEMYEGLNHLAEKVGD	403		

```

Db      347  APGRPAKVVA1PRLSPGLQ2PTMEDLAD3YK INNNGVGLVDP4RRAREL5ELGLSHLQ6ASID 406
Oy      404  GYKIDVH-----LBMUCEDYGRVDLAKAY 430
Db      407  GYKVDVHRRKWI1LKIKDNATERFWL2KEIMRLKDCVLLQ3LLEMCVEEYGRVELAKAY 466
Oy      431  YKAMKRSINKHFKGNGV1ASMEHNDMEFLGTEALISLGRVD2PWC3DPBGDPNG4FTWLO 480
Db      467  FAGLTESYRRH1NGV2ASMEHNDMEFLGTEAV3ALGRGDDPWC4DPBGDPNG5FTWLO 526
Oy      491  GCHWHVCANDSLMNGN1FHPMDMFOSTHPCAPAFHAASRAISGGPIVVS2OGKHNFDL 550
Db      527  GCHWHVCAYNSLMNGAF1HPMDMFOSTHPCAPAFHAASRAVSGGPVVS2DAVGC3HPD4DL 566
Oy      551  KKLVL1PDGSLRSEY2ALPT3BDCLFEDPLNGHETMLKIMNLNKTGYIGAFNCOGGG4WCR 610
Db      587  RRLAL1PDGSLRSEY2ALPT3BDCLFADPLHDKKMLIMVNNKRSYLGAFNCOGGG4WMSR 646
Oy      611  ETRNROCSOY1SKRYT2SKTNPKDIEMHSGENPISIECVTFALYLYOAKLILSK3PSQDL 670
Db      647  EARRMCAAGSV1TVTAASPADVEMSHGG-----GGSFRAVYFVARKLQ2LIRDESV 701
Oy      671  DIADLPEFELIY1SPYKLL--QTSLH2PAPIGLV3NMLNLTSGAIO4SD---YDDLS5SV 725
Db      702  ELTLEPFYELLVVA1PVRAIV2SPELGIGAFAPIGLANNLNGAGVAG3GEARAKGDVAA-E 760
Oy      726  IGVNGCGEMRFA1SKPRAC2IGDEEDY3GP4KDOOMV5VOV6PMPIBSSG7ISVIEFL 784
Db      761  VAVKAGAGMAV1SSARRPL2CVNGQDAEFKI-EDGIVTAD3VP--TGS4SKLSRVE5EY 816

```

	RESULT 5				
O93XK2					
ID	O93XK2	PRELIMINARY:	PRT:	853 AA.	
AC	O93XK2:				
D7	01-DEC-2001 (TREMBLrel. 19, Created)				
D7	01-DEC-2001 (TREMBLrel. 19, Last sequence update)				
D7	01-MAR-2002 (TREMBLrel. 20, Last annotation update)				
DE	Stachyose synthase (EC 2.4.1.67).				
CN	SrsI.				
OS	Pisum sativum (Garden pea).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
CC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eucosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.				
NCBI_TaxID=3688;					
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-CV. WUNDER VON KEVEDON; TISSUE=SEED;				
RA	Peterbauer T.:				
RT	"Molecular cloning of a cDNA encoding for stachyose synthase from pea				
RT	seeds."				
RL	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AJ311087; CAC38094.1; -				
KW	Glycosyltransferase; Transferase.				
SEQUENCE	853 AA; 95890 MW; CB10F18CAD37B38C CAC64;				
Query Match	45.3%; Score 1914.5; DB 10; Length 853;				
Best Local Similarity	43.7%; Pred. No. 5,2e-147;				
Matches 379; Conservative 147; Mismatches 243; Indels 99; Gaps 13;					
OY	1 MAPSKNGSGNVVSFDGLNDMSPPAIDGSDFTTVGHSLDVPENIVASPSPTYSIDK- 59				
	: : : : : :				
Db	1 MAPPLNSTTSMNLTKESIFDLSE-----RKRFVGKGFLEPHVPEENV--SPRSFSISCKP 52				
OY	60 ---SPVSV-----GCFTVGDASEPDSRHHVSYIGKLDIRFMSTFRKKVMWTTHM 105				
	: : :				
Db	53 SESNAFPSELLOKVLAYSKRKGFGFGEHETPSDRLNLSIGSFNGKDFLSIFRRKTWSW 112				
OY	VGRNGDLESFOQVILTEKKS DSGRGVFLLVEGGPPRTSIOPGDGDPVDVCVESGSSKY 165				
	: : :				
Db	113 IGKSSDIOMETOWLLIEPET-KSYVIIPITKECFNSALPGENDHKIIIAEGSGSTKY 171				

0Y	166	VDASRSMILYHACGDPAALVCEAKVYRTHLGTFFRLLEETPGIYDKFCWCMWDAFYL	225
Db	172	KESTNSTAIYAHFNSPNFLDKKEKSAIRVHLNFSRLLEETIPNLVDKRCWCMWDAFYL	231
0Y	226	TVHPQGVIEGVYRHLVDGCCPGIYVLIDDQWOSICHDSPITKEGMOQTVAGEOMPCRLK	285
Db	232	TVNPIGFIHGHLDDESKSGVEBREFYIIDGQMOSIFSDGYDNPEDAKNVLVLGGEGMSRLHR	291
0Y	286	FOENKFPDYNNPKATGR-----	304
Db	292	FDECKTFKFKYESGLLIGNSPPRYDNNFTDLLKGIEHKEIRKREELAISSKSSDLAEIE	351
0Y	305	-----AGQK-----GMAFIDELKEEFKTEVHHYVMAHLCGYW	337
Db	352	SKIKRVKEIDDLFGGEQFSSGGEKSEMSEXYGLAKFTKDLRTKFKGGLDDYVMAHLCGAW	411
0Y	358	GGLRPQVQGLDEARVLIQPVLSPGIOMTMEDLAVDKIYLNHKVGLVPRPEAKEMVEGLLAHL	397
Db	412	GGVRETEHLL-DTKIVPCPKLSPLGIDGTMEDLAVEIKASGILVHPQANDLDSMHSYL	470
0Y	398	EKVYIDGVKIDVYIHLLEMLCEDYGGRYDLAKAVYKAMTKSINKHFKNQCVIASMENHNDP	457
Db	471	AESGITGVKVDYIHSLEYVCDYGGRYDLAKVYVEGLTKIVKNFNQNGMIASMQHNDP	530
0Y	458	MFLGTEAISLCRGVDDFECTPSPGSDPNCTFMVLQCCHWKHCANDSLMNGNTIHPDMHFOS	517
Db	531	FELGTRKQISMGRVGDDEFQOPNDDPMGSFMLQSVHMHICSYNSLWGMQMIQPMDFIOS	590
0Y	518	THPCAFAHAASRAISGCPYIVSDSVGKHNFLDKRLVLPDSILRSEYVALPFTDCLEFD	577
Db	591	DHVCKAFHAGSRALQCGYIYSDVWVSHDFOLIKKLVFPDQITPKCIYFLPLPRDCLFKN	650
0Y	578	PLHNGETMLKTIWNLNKTGVTIGAFNCOGGCMCRETRRNOCSQYSKRVYTSKTNPKDIEWH	637
Db	651	PLFDEHTYLLKIMNKNKGGVIGAFNCOGAGMDPIMQFRGPEBCYKPIPGTVHTEBEMD	710
0Y	638	SGENPISIEGVKTEFLVLYQXAKKL-IISKRPQDIDLALDPEPELIVSPVKLIQTSLH	696
Db	711	OKEEIYSHLGKKEEYVYVILNQABEELSTMLKSEPIQOFTIOTSTFELYSFVYATKLC-GGik	769
0Y	697	FAPIGLVMLNTSGAIOQSDVDDDLSSVEIGVCKGCGEMRVFASKRPRACRIDEDGVFKY	756
Db	770	FAPIGLITMNFNSGCTVIDLELYVG--NGAKIVKKGGSFLAYSSSPKFKOLNCGEDVFEW	837
0Y	757	DQDQMVVYQVFWPIDSSGGSIVLEYLF 784	
Db	828	LQDGKLCVNVFM-IEEAC-GVDSMEITFP 853	

ID	Q9SBZ0	PRELIMINARY:	PRT:	857 AA.
AC	Q9SBZ0:			
DT	01-MAY-2000 (TREMBLrel_13, Created)			
DT	01-MAY-2000 (TREMBLrel_13, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel_20, Last annotation update)			
DE	Galactinol-raffinose galactosyltransferase (EC 2.4.1.67).			
OS	Phaseolus angularis (Adzuki bean) (Vigna angularis).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurosidids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.			
OX	NCBI_TaxID=3914:			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20117502; PubMed=10652123;			
RA	Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;			
RT	"Stachyose synthesis in seeds of adzuki bean (Vigna angularis):			
RT	molecular cloning and functional expression of stachyose synthase.";			
RL	Plant J. 20:509-518(1999).			
DR	EMBL: Y19024; CAB64363.1; -.			
KM	glycosyltransferase; transferase.			
SO	SEQUENCE 857 AA; 94949 MW; 85248C4B81165679 CRC64;			

Query Match 45.3%; Score 1914; DB 10; Length 857;
Best Local Similarity 43.8%; Pred. No. 5,8e-147;
Matches 370; Conservative 139; Mismatches 243; Indels 92; Gaps 10;

17 GUNDMSSPAIDGSDFTVNGHSLDVPNIYASPSPTST---DKSPSYV-----64
13 GLEPEKEVFDLSDGKLTGVGVLLSHVPEVNTF--SSFSICVPRAPPSIIQRTYTAASH 70
65 -CCFVGFDAESEDSDRHVYSGIKLKDIFMSIFRFKVMWTHVNGRNGLESETOVILE 123
71 KCGFLGFSVNSDRLINSISFCRGNFSTIFRFTWMTSTONVNGSLOMETOWILLE 130
124 KDSGRPVFLPIVEGPFRTSIOPGDDFVDCVESGSSKYVDAFSRSMYLAHGDPE 183
131 VPEI-ESYVVIPIIEKFSRSLAHGSDDHVICAESSSTQVRASSFGALVAVHAETPY 189
184 ALVKEAMKIVRNHLGTFRLLEKTPRGIVDKFGCTWDAFYLTVHPQVIEGVRLVDG 243
190 NLMREAYSALRVHLDSFRLLEKTPRIYDKFGCTWDAFYLTVPVNGVWHLGKDFSEGG 249
244 CPGGLVLDDGMSIGHOSDPTKEGNOTVAGEOMPCLLFOEYVKRFDYVNPKATGP 303
250 VAPRFVYIDGMSQVNFDEDEPNEDAKNLVLGGEQNTARLHREBQDKFRKTKQKGLLP 309
304 RA-----GQK--308
310 NAFSPNETIKELISGIEAEHLGKQAAISAGSDLAIELMYKVRREIDLEGCKG 369
309 -----GKAFIDELKEGFKTYEHHVYVHALCGYWGRLRPQVPGLEAPRY 352
370 ESNEGGCCCAAECCGMDFTTDLRTEKFGKLDVYVHALCGMGVPRGTTHL-DSKI 428
353 IOPVLSPGLOMTMEDLAVDKITLVHKGVLVPRKAEEMVGLHAHLEKVIDVYKIDVHIL 412
429 ITCRLSPGLVGTMKDLANDKIVEGSIGLVHPQANDLDSMHSYLAQVIGVTKIDVHIS 488
413 LEMLEDYGGVRLDAKAYYKAMTKSINKHFKNGVYASMEHNDPMFLGTEAISLGRVD 472
489 LEYVEEYGGVRLAEKAYYDGLTNSIKNFNGSGLIASMOGNDFFLGTQKIDPGRVD 548
473 DWCHDPSDDPNTGTMLOGCHVHCANDSLAMGNFTHPMDMFQSTHPCAAHAAIRAIS 532
549 DMFQDPPNDGPMGVFMLOGVHMHCSYNSLWMOQIOPMDMFQSDHECAKFNAGRAIC 608
533 GGPVYSDVSGKHNFDLLKLVLPDGSILRSEYVALPTRDCLFEEDLHNGEMTKMNLN 592
609 GGPVYSDVSGKHNFDLLKLVLPDGSILRSEYVALPTRDCLFEEDLHNGEMTKMNLN 668
593 KFTGVYGAFCNCGGMCRETRRNOCFSQYSKRVTSKTNPKDIEMHSGENPISIEGVKTPA 652
669 KYGVYIAGAFNCGGACMDPRGKFKGPECKYKAISCVHVTLEVEMOQKKEAEHMGKAEVY 728
653 LLYOAKKILISK-P-SQDIDLADPEFEFLTVSPYTKLIQTSLHAPRIGLYNMLTSGA 711
729 VYLNDAEYVHLMTVPSEPLQTLTOPSTFELYNFVPEKSGSSNIKFAPIGLTNMFNSGT 788
712 IGSVDYDDLSVEIGVGGCEMRVASKPRACRIDGSEVGFKKDDQDMVYVQVWPMPID 771
789 IOLELEYE--KDYKVKVKGGRFLAYSTDSPKFKQDLNGSDAAFWLDPDKLTLLNLAW-IE 845
772 SSSG 775
846 ENDG 849

RESULT 7
Q8RW08 PRELIMINARY: PRT: 863 AA.
Q8RW08
AC 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Stachyose synthase (EC 2.4.1.67).

GN STS.
OS Stachys affinis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Lamiales; Lamiaceae; Stachys.
OX NCBI_TaxId=168825;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RA Pesch M., Schmitz K.;
RT "Molecular cloning of a cDNA encoding for stachyose synthase from
RT Stachys sieboldii."
RL Submitted (AUG-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL: AJ344091; CAC66963.1; -;
KW Transferase; Glycosyltransferase.
SQ SEQUENCE 863 AA; 95227 MW; D96F666090DEPD CRC64;

Query Match 44.2%; Score 1868.5; DB 10; Length 863;
Best Local Similarity 42.3%; Pred. No. 3e-143;
Matches 366; Conservative 143; Mismatches 253; Indels 103; Gaps 9;

16 DGLNDKSP-----FALDGSFTVNGHSLDVPNIY-----ASPSPT 55
6 DPLISSIFSPDISVKKDNAMELVGKSLVKNVPLSEIPENVTFRSSICQSSGAPAPLY 65
56 SLDKSPSVSGCFVGPDPASPDSDRHVYSGIKLKDIFRMSIFRFKVMWTHVNGRNGDLES 115
66 NRAQSLNSCGFLGFSQKESADSVNSLGRFTNRVSTIFRFTWMTSTONVNGSLOSDIO 125
116 ETQIVILEKSDSGRPVFLPIVEGPFRTSIOPGDDFVDCVESGSSKYVDAFSRSMYLA 175
126 ETQWIMLNLP-EL-KSAVAVIPIVEGKFSRSLAPRGKHVHLISAGSSSTCVKTSSTFISAV 184
176 LHAQDPRFLVKEAMKIVRNHLGTFRLLEKTPRGIVDKFGCTWDAFYLTVHPQVIEG 235
185 VHVSDMPYTLMKDGYTAVAVHLDTEFLIEKSPAPLVNFKFGCTWDAFYLTVEPAGIWMG 244
236 VRLVHVGCGPGLVLLDDGMSIGHOSDPTKEGNOTVAGEOMPCLLFOEYVKRFDYVNP 295
245 VKEFSDGGSFPRFLIIDDGMSINIDGQDPNEDAKNLVLGQVOTARLHREBQDKFRK 304
296 VNPKATGP-----303
305 KGGSMGPRVYFDPKPRLLISKATEIEGVKARDKALIOSGITDLSQYEIKLKLKREL 364
304 -----RAGQKMAFIDELKEGFKTYEHHVYVHALCGYWGRL 340
365 DEMFGGNDGKSSKGCSDCSKSONSGMAFTNDLRNFKGLDIDYVWHALAGAMGV 424
341 RPOVPLPEARVYIOPVLSPLGLOMTMEDLAVDKITLVHKGVLVPRKAEEMVGLHAHLEK 400
425 KPGATHL-NAKLEPCKLSGLDGTMTDLAVVKILBESIGLVHFDQAEEDYDSMHSYLSKV 483
401 GIDGVKIDVYHLEMLECDYGGVRLDAKAYYKAMTKSINKHFKNGVYASMEHNDPMFL 460
484 GITGVKVDVYHLELVYSEVNGVELGKAYYGLSLSKNNNGSGLISSMOGNDFFLL 543
461 GTEALISLGVAGDFFWCTDPSGDPNNGTVMLOGCHVHCANDSLAMGNFTHPMDMFQSTH 520
544 GTEOISMGRVGDFFWQDNDGPMGVFMLOGVHMHCAVNSMMQGIIPDMDFQSDHC 603
521 CAAFHAASRAISGGPVIYSDVSGKHNFDLLKLVLPDGSILRSEYVALPTRDCLFEEDPLH 580
604 SAKFNAGSRAICGGPVIYSDVSLGHDHFDLLKLVFEDGTIPKCIHPLRTRDCLFENPLF 663
581 NGETMLKIMNLKFTGVYIAGFNCOGGMCRETRRNOCFSQYSKRVTSKTNPKDIEMHSGE 640
664 DSKTIKIMNFNKGYSVGAFAFCOGGMDPKQRIKGYSECKKPLSGSVHVDIENDQAV 723
641 NPISIDGVKTFALYLOAKKILSKPSQD-LDIALDPEFEFLTVSPYTKLIQTSLHAP 699
724 EATKMGAEVAVYVLTSEKLLTLTPESDPIPTFTKSTFEIFSVPKIKLGO-GVKFAP 782


```

Db 125 LPLEGQFRTVLQGNDDQLQICIESGKAVETEGQNMNNVYHAGTNPEDTITQAVKAVE 184
Oy 195 THLGFRLLEEKTPPGIVDKFGMCTDAFYLTGVHPOGVIEGVHNLVDGCGPCLVLIDG 254
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 185 KHTQFTHHREKTPVSEVMEFGCTDAFYTDVTAAGVAGGLSLSLGEGAPREFLIIDG 244
Oy 255 MOSIGHD--DPTKEGNNQTVAGEOMPRLKFOENYKFRDYVNPKATGPRAQOGKMA 312
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 245 WQOIGSEKNDP---GV-AVQGAQFASRLTGIKRENTKFOSEHNEET-----GGLKR 293
Oy 313 FIDELKGEKFEYEHVYVHNLGCGYNGGLRPQVGPL--EARYIQVLSGLQMTMEDLAY 370
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 294 LVDETRKE-HGVKSVYVHMAAGVGVKPSAAGMEHYEPALAYPQASPGVGNOPDIY 352
Oy 371 DKIVLHVGLVPEPEKAEMEYEGHLEKVGIDGVKIDVHLEMLCEDYGGVLDLAKAY 430
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 353 DLSVGLGGLVHPRVHRRYDELHATLACGVGVKVDQNIYETIGAGHGVALTRAY 412
Oy 431 YKAMTKSINKHKGNGVIA SMEHCNDFMFLGTETALSLGRVGDDEFWCTDPSGDPNGTFWLQ 490
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 413 HALEASVARNFPDNCISCMCHNTDMLY-SAKQTAIVRASDPFYPRDPAS-----462
Oy 491 GCHMYH---CANDSLMGNFIHPDMDMFQSTHPCAFHNASRAISGPTIYSDSVGKHN 547
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 463 --HTVHISVAANTLFLGSEFMQPDMDMFHSLHPAAEYHGAARAIGCPTIYSDKPNHNE 520
Oy 548 DLKLKLVLPDGSILRSEYVALPTROCLFEDPLHNGETMLKINLKFYIGAFNCGG 607
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 521 DLLRKVLVLPDGSVLRNQLRGRTROCLFSDPARADGSLKINMMNKACAGVGVFNCGAG 580
Oy 608 KCRETNRNOCFSQYSKRYVSKTNPKDIE-----WHSGENPISIEGVKTFALYLYO 657
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 581 MCRVAKRTIHDEAPFTLTGSVAEDVEAIAQAGTGM-GGE-----AVVYARH 629
Oy 658 AKKLISKPSQDLIDLDPFEFELITVSPYTKLIQTSLHPARTGLVNLMTSQAISQV 717
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 630 ADELVLPRGATLPTLKLLEYELFHVCPV-RAVAPGVSAFPTGLLHMFVAGAVECTV 688
Oy 718 ---DDDLSSVEIGKCGEMRFAKPRACRIDGEGVFYKYPDO-MVVYQVPM 769
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 689 ETGEDNNAVGLRVRCGGRFAGVCSRRPAKCSVDSADVEFTYDSIGLTVADYVVP 744

```

RESULT 10

```

OQ3408      PRELIMINARY:      PRT:      765 AA.
AC 043408:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Putative Imbibition protein.
OS Brassica oleracea (Cauliflower).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Brassica.
OX NCBI_TaxID=3712;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. ALPHA BALOMA. TISSUE=ROOT TIPS.
RA Fujikura Y., Karsen C.K.;
RT "Cauliflower cDNA encoding a putative imbibition protein.";
RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL; X79330; CAA55893.1; -.
SQ SEQUENCE 765 AA; 84084 MW; 73398603048E9B58 CRC64;

```

Query Match 34.9%; Score 1473.5; DB 10; Length 765;
 Best Local Similarity 37.6%; Pred. No. 3.9e-111;
 Matches 297; Conservative 155; Mismatches 253; Indels 85; Gaps 18;

```

Oy 21 MSSPFIIDSDFVNGHSFLSDVENIVASPSPTSIDKSPVSGCVGFADASEPDSRHV 80
    :||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 3 ITSNISVQNDNLVVGOKTILTKIPDNILITPVAGASDS-----GAFIGATFRQSKSLHV 57

```

```

Oy 81 VSIGLKLIREMSIFRFKYMWTTHVNGRNGDLESETQIVLEK-----SDSGRPYVFL 134
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 58 FPIGVLEGLRPMFCFRFKLMMMTQRMGASGKQIPLETFQFWLIESKDEVDNDAPVYTVF 117
Oy 135 LPVIEGPRFTSIOPGDDVDVDCVESGSKYVDASTRSMILYLAAGDPPALVYEAUKIYR 194
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 118 LPLEGQFRTVLQGNKNEIEICIESGKAVETEGQNTHLVYHAGTNPEDTITQAVKAVE 177
Oy 195 THLGFRLLEEKTPPGIVDKFGMCTDAFYLTGVHPOGVIEGVHNLVDGCGPCLVLIDG 254
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 178 ROMQFTHHREKTPVSEVMEFGCTDAFYTDVTAAGVAGGLSLSLGEGAPREFLIIDG 227
Oy 255 MOSIGH--DSDPTKEGNNQTVAGEOMPRLKFOENYKFRDYVNPKATGPRAQOGKMA 311
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 238 WQOIGSEKNDP---GV-AVQGAQFASRLTGIKRENTKFOSEHNEET-----GGLKR 293
Oy 313 FIDELKGEKFEYEHVYVHNLGCGYNGGLRPQVGPL--EARYIQVLSGLQMTMEDLAY 370
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 286 SVYDNMK-QRHNVKQYVAMIALAGVGVKPRASGMEHYDSALAYPQASPGVGNOPDIY 344
Oy 370 VDKIVLHVGLVPEPEKAEMEYEGHLEKVGIDGVKIDVHLEMLCEDYGGVLDLAKAY 429
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 345 MDLAVHGLGLVPPKAVFNFYNELHSTLASCGIDGVKVDQNIYETIGAGHGVALTRAY 404
Oy 430 YKAMTKSINKHKGNGVIA SMEHCNDFMFLGTETALSLGRVGDDEFWCTDPSGDPNGTFWL 489
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 405 YHQAELASVARNFPDNCISCMCHNTDMLY-SAKQTAIVRASDPFYPRDPAS-----462
Oy 490 GCHMYH---CANDSLMGNFIHPDMDMFQSTHPCAFHNASRAISGPTIYSDSVGKHN 546
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 456 --HTVHISVAANTLFLGSEFMQPDMDMFHSLHPTEYHGAARAIGCPTIYSDKPNHNE 512
Oy 547 FDLKLKLVLPDGSILRSEYVALPTROCLFEDPLHNGETMLKINLKFYIGAFNCGG 606
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 513 FDLKLKLVLPDGSVLRNQLRGRTROCLFADPARDGTSLKINMMNKFTGIVGFNCGA 572
Oy 607 GWCETNRNOCFSQYSKRYVSKTNPKDI-----EWHSGENPISIEGVKTFALYLYO 657
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 573 GWCMDTKKRNHDTSPGTGLGLVARBADADLISEVAGDM-GGDS-----IYAYAK 621
Oy 658 AKKLISKPSQDLIDLDPFEFELITVSPYTKLIQTSLHPARTGLVNLMTSQAISQV-- 715
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 622 SGEVLRLPRGASIPTLKLLEYELFHVCPV-KEITASISFAPITGLDMFNSSQAISMET 680
Oy 716 ---DYDDLS-----VEIGKCGEMRFAKPRACRIDGEGVFYKYPDO 758
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 681 NVTYDEKPELSSSVSENRSPALISLGVKCGRGAVSSQRLRCANDGTETERNYDA 740
Oy 759 D-QMVVYQV 767
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 741 EVGLVTLNLP 750

```

RESULT 11

```

OQ94A08      PRELIMINARY:      PRT:      773 AA.
AC 094A08:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Putative Imbibition protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC YAMADA K., LIU S.X., PHAM P.K., BANH J., BANNO F., DALE J.M.,
RA Goldsmith A.D., JIANG P.X., LEE J.M., ONODERA C.S., QUACH H.L.,
RA Tang C., TORIUMI M., YAMAMURA Y., YU G., YU S., BOSEER L.,
RA Carlini P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,

```


OY 711 AIOQVD-----YDDLS-----VEIGKCGEMKRVFASKKPRAC 745
 DB 676 AIESIDINHTOKNPEFFGEISSASPALSDNRSPTALVSYSVRGGCRGAYSSQAPLKC 745
 OY 746 RIDGEDVGFRIYDOD-QMYYVQVP 767
 DB 736 AVESTETDETYDAEVLVTLNLP 758

RESULT 13
 OY 08RX87 PRELIMINARY: PRT: 844 AA.
 AC 08RX87;
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE AT5g20250/ES024_140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsie.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shinn P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamita A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Soultmaki A., Tang C.C., Toriumi K., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RT Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY090237; AAL90901.1; - ZAC8AB0EA43F8056 CRC64;
 SQ SEQUENCE 844 AA: 94215 MW: 2AC8AB0EA43F8056 CRC64;

Query Match 34.2%; Score 1446.5; DB 10; Length 844;
 Best Local Similarity 39.5%; Pred. No. 7.2e-109;
 Matches 305; Conservative 140; Mismatches 255; Indels 73; Gaps 18;

OY 20 DMSPPAIDSD--FTVNGHSLSDVPENIVASPSPTSIDKSPVSGCFVGDASEPDS 77
 DB 95 DMTIKPAVRISDGNLIKRTILITGVPDVNTT-----SASEAGPVE-GYFVGAVFNKEES 149
 OY 78 RHVYSIGKLDIRFMSIFRKVMWTHHWGRNGDLESETOIILEKSDSG----- 128
 DB 150 KHLVPIGLTILNSRFMSCFRKLMMMAORMGEMGRDIPYETQFLLVESNDGSHLESOGANG 209
 OY 129 -----RPVYFLRPIVEGPFRTSIQPGDDFVDVCEVSGSSKVVVDASFRLYLHAGDDPF 183
 DB 210 VECNQVYVYFELPLIEGSPFSCOGNVNDEVELCLESQDVTKRSTHSLYLHAGDDPF 269
 OY 184 ALVKEAMKIVRHLGTRFLLEKTPRGIVDKFGMCTWDAFYLVNPGVIEGVRYHLVDG 243
 DB 270 QITTDIAIRTKLILNSFRHREKRLPGIYDYFGMCTWDAFYEQVTOEGVAGLKLASLAG 329
 OY 244 CPGGLVLIDCGNSIGHDSRPIRKEGMNQV-AGEQMP-----CRLLKFGQNKYFRDQVNP 258
 DB 330 TPKEFYIIDGWSVERDA-----TVEAGDEKKESSPIFLGLGKENEKFK----- 374
 OY 299 KATGPAGOGKMAFIDELGKFEKTEVHYVWMLCGYMGGLRPQVYRLEAPRIQ-PVL 357
 DB 375 KKDPPNVIKNIYIKIKEKKG-----LKYVYVWMLAITGYMGVNR---PGEYGVSMKPYRM 427
 OY 358 SPGLQMTMEDLAVDKIVLHKVGLVPEKAEEMV EGLHAELEKVIDGVKIDVILHEMLC 417
 DB 428 SKGVENDPTWKTDMVTIQLGLVSPKKYKFFYNELSHYLADAGVGDVKKVQDCVLETTIG 487
 OY 418 EDVGGVNDLAKAYKAMTKSINKHFKGNGVYIAEMHENDMFLEKTAISLGRGDDPFMCT 477
 DB 488 GGLGGRVELTROFHALDSSVANKFPDNGCIACMSHTDALTALYCSKQAAVI-RASDDPYPR 546

OY 478 DPSGDNGTFMLOGCHMVHCAN---DSLWGNFIHPDMDFOSTHPCAAFPNAAISGC 534
 DB 547 DPVS-----HTIHASVAYNSVFLGERFMQDMDMFHSHHAAEVAHSAARAIISGC 595
 OY 535 PIYSDSVGKHNFDLKKLVLPDGSILRSEYVALPFRDCLFEDPLHNGETMLKIMLNKF 594
 DB 596 PLVSDSPCKHNFELRLKLVLPDGSILRLRALPGRPTKDLFADPARDGVSLKIMWNKKY 655
 OY 595 TGVIGAFNCOGGWCMETRRNOCFSQYSKRVYKTPNPKDIEHS-----GNPISIEGVKT 650
 DB 656 TGVIGVYNCOGAAMSSSTERKNIFHQRTDLSIGRQV--HSISEASTDPTTWNG--D 711
 OY 651 FALYQAKKILSKSPQDIDIALDPFEELIVSPYTKLIQTSLHFAPIGLVNMINTSG 710
 DB 712 CAYSSOSRGLIYMPYNSLPLVSLPKLREHIEFTVSPISHLVD-CYSFAPIGLVNMINSOG 770
 OY 711 AIOQVYDDDLSSVEIGVKGCGEMRVFASKKPRACRIDGEDVGFYDQGMV 763
 DB 771 AIGLRYEAEKMKVMEVKCGKFGSYSVKPRCVESNEIAFEYDSSGLV 823

RESULT 14
 OY 09M4M7 PRELIMINARY: PRT: 779 AA.
 AC 09M4M7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Putative seed imbibition protein.
 GN SIP.
 OS Persea americana (Avocado).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
 OX NCBI_TaxID=3435;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN-CV. HASS; TISSUE-MESOCARP;
 RA Zamorano J.P., Evans A.D., Doplco B., Lowe A.L., Wilson I.D.,
 RA Merodio C., Grierson D.;
 RT "Isolation and characterization of cDNAs for mRNAs regulated during
 RT cold storage of avocado (Persea americana Mill.) fruit."
 RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AJ133148; CAB77245.1; -
 SQ SEQUENCE 779 AA: 85368 MW: C3A8BA43160316785 CRC64;

Query Match 33.5%; Score 1413.5; DB 10; Length 779;
 Best Local Similarity 36.7%; Pred. No. 3.1e-106;
 Matches 293; Conservative 144; Mismatches 268; Indels 93; Gaps 17;

OY 26 AIGSDFTVNGHSLSDVPENIVASPSPTSIDKSPVSGCFVGFDAEPPDSRHVYISGK 85
 DB 8 SINDGNLVYNGKTIILGVDPNIVILTRPTGDG-----VACFEGATASESESHVFPMTG 62
 OY 86 LKDIRMSIFRFVWMTTHVWGRNGDLESETOIILEKSDSGRP-----YVFL 134
 DB 63 LEGIRFCCRFKLMWMTQMGKGRDVPLETOFMLESKGAALIDDDEEAPTYIVTF 122
 OY 135 LPVIEGPFRTSIQPGDDFVDVCEVSGSSKVVVDASFRLYLHAGDDPFRLVKEAMKIVR 194
 DB 123 LPLEGOFRVAVLOGNSNOIEILLESDDCAVPTNQGMYLVYHMGANPFRVINDAVKAVE 182
 OY 195 THIGTRFLLEKTPRGIVDKFGMCTWDAFYLVNPGVIEGVRYHLVDGGRPGVLVLDG 254
 DB 183 KHLXSOHLEKKAIRPSLIDFMGCTWDAFPTDYTDEVEEGSLKSLSGGTRPRRLIIDDG 242
 OY 255 WOSIGHSDSDITRKEGMNQVA-GEOMPCRLKFOENYKFRDYVNPAPKATGPAGOGKAKAR 313
 DB 243 WQDIGSE--TRDSDNCVYXEGQAFASRLTGTIKENDK-----QKNGKSEHVPLGLKLV 293
 OY 314 IDELGKFEKTEVHYVWMLCGYMGGLRPQVRLP--EAAYIQPVLSPLGLQMTMEDLAVD 371
 DB 294 VDAK-QHNHVKEVYVWMLAGYMGVKKPPAAGMEHYDTALAYPVOSPGVWGNOPDIVMD 352

```

OY 372 KVLKLVGAVPEPERKEAEMYECEGHAEHLAEVVDGVKIDVIHILEMCEDEYGGGRVDAIAY 43
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 353 SLISVAGLGLVHPRKFKFNXNNEHLALXSCGVNGVAVDQNIETILGAGHGGRSVLSLTSYI 41
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 432 KAMTYSINKHEFGANGVIAAMEHCNDMFELGTFAISLGRVGDDFWCCTDPSPGDNFTWLOG 49
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 413 QALESSIRARNEFPDNCICACMCCHNTDSIY-SAKQFVAIVRASPDVEPRDPAS----- 46
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 442 CHMYH----CANDSLMGAFIHEDMDMPGSTPHPCAFHAASAFAISGCPITYSDSGKHNF 54
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 462 -HTHVSSVANNSLFLGEFMOPDMCMFESLHPAAEYHGAARAVGCCPIYVBDKPFHNFE 520
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 549 LLLKLVLDPDGSILRSEXYALPTRODLFEFDPLHNGETMTKLNNLKFTVIGAFNCGCGW 608
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 521 LLKLVLPLDGSTVLRARLRGRTROSLFDPDPAADVSLIKIMNMKKCLGAVGFNCGAGW 588
      |||||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 609 CRETRRNOCFSOYSKRVTSKTNPKDIE-----WHSGENPISIEGWTFALTYLOAK 655
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 581 CKITKTRTIHDAAPGTLTGSIHRAHDVEFINOLAGDW-NGE-----VIFVTYGSG 622
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 660 KLILSPSODDIALDPPREFELITYSPYTKLIQTSLHRAPIGLVMNLTSQAISOVD--- 711
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 630 EVWLRLPKCASTPVLTLEVLDELXHCIP-KETSNISAPJGLDLMFMSGAGVAGQFDVRM 688
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 717 -----YDDDLSS-----VEIQKCKCEMRVAFASKRPACRIDEDV 752
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 689 DSNNAEPPLFPGKVASKLSSLSNNGOSSAIVYLVRRCGFAGFVASORPLKCTVDLVET 744
      ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
OY 753 GFKYIQ-DQMYYVOVPWP 769
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 749 EFNYDSVTGLVTLIIIPV 766
      :|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

RESULT 15
O9LFZ7 PRELIMINARY; PRT: 1170 AA.
AC O9LFZ7;
Dt 01-OCT-2000 (TREMBLrel. 15, Created)
Dt 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
Dt 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
De F20N2.14.
Os Arabidopsis thaliana (Mouse-ear cress).
Cc Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Cc Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidaeae;
Cc OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
Cx NCBI_TaxID=3702;
[1]
Rp SEQUENCE FROM N.A.
Ra Khan S., Brooks S., Buehler E., Chao Q., Johnson-Hopson C., Kim C.,
Ra Shinn P., Altafi H., Bei Q., Chiu C., Chlou J., Choi E., Conn L.,
Ra Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
Ra Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,
Ra Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thavetl A.,
Ra Totiumm M., Vaysberg M., Yu G., Federspiel W.A., Theologis A.,
Ra Ecker J.R.;
Rt "Genomic sequence for Arabidopsis thaliana BAC F20N2 from chromosome
Rt I."
Rt Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
Rn [2]
Rp SEQUENCE FROM N.A.
Ra Ecker J.R.;
Rt Submitted (JUL-1997) to the EMBL/GenBank/DDBJ databases.
Rn [3]
Rp SEQUENCE FROM N.A.
Ra Ecker J.R.;
Rt Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
Rn [4]
Rp SEQUENCE FROM N.A.
Ra Shinn P.P., Buehler E.E., Dunn P.P., Feng J.J., Kim C.C., Li Y.Y.,
Ra Walker M.M., Altafi H.H., Araujo R.R., Conn L.L., Conway A.A.B.,
Ra Gonzalez A.A., Hansen N.N.F., Hutzel L.L., Kremeretskala I.I.,
Ra Lenz C.C., Li J.J., Liu S.S., Luo S.S., Rowley D.D., Schwartz J.J.,
```

RA Torilumi M.M., Vysotskaia V.V., Yu G.S., Davis R.W.,
 RA Federspiel N.N.A., Theologos A.A., Ecker J.J.R.;
 RL Submitted (Dec-1998) to the EMBL/Genbank/DBJ databases.
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (May-2000) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bai B., Chin C., Chou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Makharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thavert A., Torilumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologos A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AC0082338; AAF79504.1; -
 DR InterPro; IPR004798; Cax; 1.
 DR TIGRFAMs; TIGR00378; cax; 1.
 SQ SEQUENCE 1170 AA; 129057 MW; 98B43A04E3F66D4 CRC64;

Query Match 27.0%; Score 1140; DB 10; Length 1170;
 Best Local Similarity 33.3%; Pred. No. 1.le-83;
 Matches 259; Conservative 135; Mismatches 208; Indels 176; Gaps 20.

QY	26	AIDSDPFWNHSFISDPPENIVAS--SPYISIRKSPVSGCFGFPASBDSRHVYSIG	84
QY	26	AIIDSDPFWNHSFISDPPENIVAS--SPYISIRKSPVSGCFGFPASBDSRHVYSIG	84
Db	8	SVTSDBLVVLGHRLVGHVPERVLVTPASGNALID-----GAFIGVNDQGSHRVPSLG	61
QY	85	KLKDRFMSITRFVWVWVTHHVGKNGDLEBETQVILEKS-----DSGRVYVLLP	136
Db	62	KLEDRFVCFVFRKILMWTOFMKNGKEIPCEPTFLVEAOGSDLCGRQSSSYVFLP	121
QY	137	IVEGFRTSIPRGDDDFDVCVESSGSKVVDASRSMILYLHAGDPPALVKEAKIY---	193
QY	122	ILEGFRVVLQGNENANELEIOLLEGSDPTVODFESHLVFAAGSDPPDVITKAKVYML	181
Db	194	RTNHLST-----FRLBEKTPRGIVDVFQWCTMDAFVLVYHPOGVLEGVRHLVD-----	241
QY	182	KSOLEKTSLCRPFIRM-----RDMIMNFGWCTWQAFYTNVTAKVQKGLSNCDLTPKAL	235
QY	242	-----GGCPRGLVLLDDGWSIGHSDDPTIKEGMNQVAGEOMPCLRLKFOENYKFD	294
Db	236	ILCSLKAGVYPRKYVYIIDGQWSQMDETSVEFMADN-----AANFANRLTHIKENHFFQ-	290
QY	295	YVNPKATGCRAGQOKMKA-----FIDELKGEFTVEHVYVMAALCGYSGSLRPOVP	345
Db	291	-----KDKGEGHRYDDPSLSLGHVITFDIKSN--NSLKVVYVWHAITGWSGVKPGVS	340
QY	346	GLP---EAVVIOPVLSPGLOMTEBLADKIVLHGVGVLPREKAEEMEGJAHNLEKVID	403
Db	341	GMEHYESVAVPVSSPGWSSSNCGLSESTIKMGLVNPKEKPFSTYNDLHSTVASYVD	400
QY	404	GVRIDVILHLEMLCEDVGYGRDLAKAYYKANTKSINKHFGNGVYIASMEHCNFMFYGTE	463
Db	401	GVKYDVQNIILETFLGAGHGGRVYKLAKKHQALEASISNPNPDNGIISCMSHNTDGLY-SAK	459
QY	464	AISLGRGDDFWCTDPSGDPMGTWLOGCMHVCAN--DSLWNGNTIHPDWMDFOSTHP	520
Db	460	KTAIVIRASDDMPWDPPAS-----PTHIASVAYNUTLFLGFMOPDMDMFLSHLP	508
QY	521	CAAFHAASRAISGGPRIVYSDSGKHNFDLKLKVLPGGSIIRSEYVALPRDCLFEEDPLH	580
Db	509	MAEYHAAARAVGCAITYSDKPRGHDNRLKRLVLRGOSILIRAKLBPRLPRELY----	563
QY	581	NGETMLKIMWLNKFTGVIGAFNCGGGWCRETRNOCFSOYSKRVITSKITNPKDIEMHSGE	640
Db	564	-----LPKD-----	567
QY	641	NPISIEGVKTFALLYQAKKILILSPRODDIALDPREFELIYVSPYTKLITQNSLHNPRI	700
Db	568	-----TSPLVTLMPREYEVFTVAVPKESDGS-KRVP	599

Fri Apr 4 12:31:33 2003

us-09-675-208-5.rspt

Page 11

Qy 701 GLVNMINTSGAITSVDYDDLLS--VEIGVKGGCGEMRVS-KKPACRIDEDVGFK 755
||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 600 GLMEFMNSGAIVSLRYDDEGTKFVVRMKLRGSGGLGVGVSSVRRPRPSVTVDSDVVEYR 657

Search completed: April 4, 2003, 11:12:35
Job time : 48 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:09:06 ; Search time 48 Seconds

(without alignments)
1570.196 Million cell updates/sec

Title: US-09-675-208-5

Perfect score: 4225

Sequence: 1 MAPSFKNNGSNVSEFDGLND.....QVWPIDSSSGGISVIEYLF 784

Scoring table: BLOSUM62

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1853.5	43.9	807	2 C65025	hypothetical prote
2	1482	35.1	757	2 S27762	Sid1 protein - bar
3	1473.5	34.9	765	2 S45033	probable imbibitio
4	1458.5	34.5	773	2 T46188	Imbibition protein
5	1140	27.0	1170	2 C96599	protein F20N2.14 f
6	490.5	11.6	357	2 T09530	probable seed inh
7	471.5	11.2	371	2 T01717	hypothetical prote
8	365	8.6	649	2 D90496	hypothetical prote
9	121.5	2.9	538	2 S76481	hypothetical prote
10	115.5	2.7	902	2 T41051	beta transducin -
11	114	2.7	623	2 T42245	probable polypepti
12	113.5	2.7	632	2 S69702	hypothetical prote
13	105	2.5	730	1 T64118	1,4-alpha-glucan b
14	103	2.4	3944	2 T19997	hypothetical prote
15	102.5	2.4	544	2 G90524	hypothetical prote
16	102.5	2.4	564	2 HMI17	hemagglutinin prec
17	102.5	2.4	582	2 S74819	extracellular solu
18	102.5	2.4	1093	2 T51503	valine-tRNA ligase
19	102	2.4	679	2 G71615	phospholipase A2-1
20	101.5	2.4	478	2 S28222	mannose-1-phosphat
21	101.5	2.4	626	2 T42246	polypeptide N-acet
22	101	2.4	514	2 A36793	hypothetical prote
23	101	2.4	1195	2 T50226	hypothetical prote
24	100.5	2.4	712	2 B96746	probable kinesin T
25	100.5	2.4	520	2 G02512	interleukin-1 rece
26	100	2.4	722	2 S27197	hydroxymethylgluta
27	100	2.4	1550	2 S60228	glutamate synthase
28	99	2.3	770	2 S76095	hypothetical prote
29	99	2.3	979	2 G90459	formate dehydrogen

30	99	2.3	1137	2 T18625	atrial natriuretic
31	98	2.3	306	2 D69753	viomycin phosphor
32	97.5	2.3	624	2 T42247	polypeptide N-acet
33	97	2.3	520	2 A25332	hydroxymethylgluta
34	97	2.3	520	2 S12736	hydroxymethylgluta
35	97	2.3	859	1 VCLJ22	env polyprotein
36	97	2.3	1002	2 T09438	toxR-activated gen
37	97	2.3	1013	2 B82276	genome polyprotein
38	97	2.3	3061	1 JN0545	hypothetical prote
39	96.5	2.3	444	2 S48408	DNA polymerase B1
40	96.5	2.3	901	2 E84210	phospholipase C (E
41	96.5	2.3	1095	1 A31225	allantoicase (EC 3
42	96	2.3	460	1 S48489	hydroxymethylgluta
43	96	2.3	520	2 S45497	1,4-alpha-glucan b
44	96	2.3	737	2 C81724	AAA family ATPase
45	96	2.3	769	2 E90158	

ALIGNMENTS

RESULT 1

C65025 hypothetical protein AT4G01970 (imported) - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: C65025

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A65001; MUID:20083488; PMID:10617198

A:Accession: C65025

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-807 <STO>

A:Cross-References: GB:NC_001268; MID:97268581; PIDN:CAB80690.1; GSPDB:GN00140

A:Gene: AT4G01970

A:Map position: 4

Query Match	43.9%	Score 1853.5;	DB 2;	Length 807;
Best Local Similarity	44.5%	Pred. No. 1.5e-138;		
Matches 366;	Conservative 146;	Mismatches 257;	Indels 53;	Gaps 14;
QY	1	MAPSFKNNGSNVSEFDGLND--SSPFAI-----DGSDFTVNGHSFLSDY	43	
DB	1	MAPLHE-----SLSSINDVIESKPLFVPIITKPIIQPNFNLSEGLCAKDSPTPLFDV	53	
QY	44	PENIVASP-SPYTSIDKSPSV-----GCFVGFDPASEPDSRHVVSIGKLDIREM	92	
DB	54	PQNVFTPTPSSHISISIDAPLPIILRRQANAHKGGFLGFTKESPSDGLTNSLGFEDREFL	113	
QY	93	SIFRFKVMVTTIMVGNRGSDLESEIOVILE--KDSGRPVYFLPLIVGPEPTSIQOPD	150	
DB	114	SIFRFKMMVSTWIGKSGSDLOAETQWMLKIPEDIS---YVAIIPTIGAFRASTLPGE	170	
QY	151	DDGVDDVCGSSGSKVVDASFRSMLYLHAGDDPPLALYKEAMKIVRTPLGFRFLIEKTPPG	210	
DB	171	KGNVLLCAISGSKTKVKESSFKSTAYTHICDNPMLKKEAPSAIRVAMNFKFLIEKKLPK	230	
QY	211	IVDKFCMCTMDAFYLVHPOGVIEGRHLVDGCGPGLVLIDGQMSIGHDSDPPIKCEM	270	
DB	231	IVDKFCMCTMDACYLVDPATITWGVKEEDGCVCKFVYIIDDQMSINPDDGELDKME	290	
QY	271	NOTVAGEQMPRLKQOENYKTRDYVNPRAKATGPRAGQKMAFIDELKCEFTVEHYVW	330	
DB	291	NVLTGGEQMTARLTSEKCKKFRNYKEESLGSDDVSGSMAAFTKDLRLRFRSISLDIYW	350	
QY	331	HALCGVWGLRPQVPLPARVYQPVLSPLGLQMTMDLAVDKIVLHKVGLVPEKAEEMV	390	
DB	351	HALCGVWGLRPQVPLPARVYQPVLSPLGLQMTMDLAVDKIVLHKVGLVPEKAEEMV	410	
QY	391	EGILHLEKRVGIDGVKIDVILHLEMLCEDYGGRVDLAKAVYKAMTKSINKHFGKNGVIA	450	

```

      : : : : | | : | | | | | : : : : | | | | | : : : : | | : | | | |
Db 411 DSMHSLASGVYGAIKIDVFOITLESLEBHGKGVLELAKAYDGLTESMINKFNSTIVIAS 470
OY 451 MENCNPFMLGTFAISLIGRGDDEFWCTDPSGDNNGTGMLOGCHMYCANDSLMGNFIRP 510
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 471 MOOCNFEFLATKQISIGRGDDEFWMODPGDQGVYMLQGVMTICSTINIMGMQIOP 530
OY 511 DMDMFQSTHPCAFHAASRAISGPIYVSDSVK--HNFDLKLKLVLPDGLSRSEYAL 568
      | | | | | | : | : | | | | | : | : | | | | | : | : | | | | |
Db 531 DMDMFQSDHVCAYHAASRAICGGPYLSDHLGKASHNFDLKLKLPDGLSRSEYAL 590
OY 569 PPRDCLFEDPLHNGETMLKIMLNKFTGYGAFNCGGCKRFRNQCQSQSKRYTSK 628
      | | | | | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 591 PPRDSLFKNPLFDEKESILKIFNFKGAGVIGTPNCCAGASPEBHRKGYKECYTTSVT 650
OY 629 TNPDKLEWMSGEPISIEGVKTFPA---LYLYQAK-LIISKPSQDLDLIDPEEFLIT 683
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 651 VHVSDLEW--DQNPBAAGSVYITGDIYLYKQOSEELIFNNSKEAMKITLFEASDLS 708
OY 684 VSPYTKLQTSLSHPADIGLVNMLNTSGAIOQSDYDDLSSEIGVKGCGEMRVASKRPR 743
      | | | | : : : | : | : | : | : | : | : | : | : | : | : | :
Db 709 FVPTVELVSSGVAFADLGLINMENCYGVQDMKVTGD-NSIRVDVKGEGRFMAVSSAPV 767
OY 744 ACIDDEEDVGFKKYDQO-MVYVQVPPRIDSSSGGISVIEYLF 784
      | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : :
Db 768 KCYLNDKEAFKWEETGKLSFFVWV--EESGGISLSLSTF 807

```

102(6) clm 34

RESULT 2
S27762
Slp1 protein - barley
C:Species: Hordeum vulgare (barley)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27762
R:Heck, G.R.; Dorsett, C.; Ho, T.H.
submitted to the EMBL Data Library, February 1992
A:Description: Cloning and characterization of a gene, Slp1, associated with seed imbibition
A:Reference number: S27762
A:Accession: S27762
A:Molecule type: DNA
A:Residues: 1-757 <HEC>
A:Cross-references: EMBL:M77475; NID:g167099; PID:g167100
C:Genetics:
A:introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match 35.1%; Score 1482; DB 2; Length 757;
Best local similarity 39.4%; Pred. No. 4e-109;
Matches 306; Conservative 128; Mismatches 266; Indels 76; Gaps 18;

```

OY 28 DGSDFTVNGHSLSDVPEINIVASPSPTYSIDKSPVSGCFGFDASEPDSRHVYSGIKL 87
      | | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 11 DGR-LAVRGRITVLSGVDVNTAAHAGAGLVN-----GAFVGATAAEAKSHHFTGTLR 64
OY 88 DLRPMALFFPKVWMTTHWGRNGDLESEIQIYLE-----KSDSGRYVFL 134
      | | | | : | | | | : | : | : | : | : | : | : | : | : | : |
Db 65 DCFEMKLFNFKKLMWMTQRMGTSGRDVPLEQFLILEVPAAGDDSDSDGDEPVLVLM 124
OY 135 LPIVEGPFRTSIOPGDDDEVDCVESGSSKVVVDASFRLMUTLHAGDDPFLVLEAKIVR 194
      | : : | | | : | : : | : | : | : | : | : | : | : | : | : | :
Db 125 LPLLEOFRVAVLQGNODELQICIESGDKAVETEOGMNWNVYIAGINPFTITQAVKAVE 184
OY 195 THLGFRLLLEKTRPGIIVDKFGMCTWDARFLVIVNPOGVLEGVNHVLDGCGRPLVILIDG 254
      | | : : | : | : | : | | | : | : | : | : | : | : | : | : | : |
Db 185 KHTQGTNHNREKKTVPFVDFWFGMCTWDARFLVIVNPOGVLEGVNHVLDGCGRPLVILIDG 244
OY 255 WQSIGDS--DPTIKGAMQTYAGDOMPCRLKLFQENYKFRDYVNPKATGRPAQOKGMA 312
      | | : : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 245 WQOIGSENDDP---GV-AVDEGAQFASRLTGIIRNTKFRQSEHNEET-----PGLKR 293
OY 313 FIDELGKFEKTVENHYVNHALGCGWGLRPQVGLP--EARVITQPLSPGLQMTMEDLAV 370
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 294 LVDEITKKE-HGKYSVYVNHAMAGYWGKVPKSAAGMEHYEPALATYPQVSGVLTGNQPDIVM 352

```

```

OY 371 DKIVLHKGVLPEPKAEEMEGSLAHLEKVGIDGVKIDVTHLLEMLCEDYGGFVNDLAKAV 430
      | : : | | | : | : | : | : | : | : | : | : | : | : | : | : |
Db 353 DSLVSLGLGLVHRRVHREYDELHAYLAACGVGVQVQVONIVETLIGAGCHVALTRAV 412
OY 431 YKAMTSINKHFKGNQVYASMEHCNDENMELGTEALISLGRGDQPMCTDPSGDNNGTFLW 490
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 413 HRALEASVARNFPDNCISQCHNTDMLY-SAKOTLAVRRASDPFPRDPS----- 462
OY 491 GCMVH---CANDSLMGNFIHPDWMFOSTHPCAFHAASRAISGPIYVSDSVKKNF 547
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 463 --HTVHSSVAYNTLFLGERMOPDMDMFHSLHPAEVHGAAARIAGGPIYVSDPKGNHF 520
OY 548 DLKLKLVLPDGLSLRSBYALPTPDLFEDPLHNGETMLKIMLNKFTGYGAFNCGG 607
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 521 DLLRKLVLPGSVLIRAPQLPGRPTDCLFSDPARDASLKLIMMNNKAGVGVFNQAG 580
OY 608 WCRETRNRQCFSQSKRYNKSINPKDIE-----MHSGENPISIEGVKIFALYLO 657
      | | : : : : : : : | : : | : | : | : | : | : | : | : | :
Db 581 WCRVAKKTRIHDEAPGTLTGSVRAEDVEALQAAGTGM-GGE-----AVYIAHR 629
OY 658 AKRLILSKPSQDLDLIDPEEFLITVSPYTKLQTSLSHPADIGLVNMLNTSGAIOQSDY 717
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 630 AGEVLRLPRGATLPTLKLREYELFHYCPV-RAVAGVSVAPFGLHMFNAGAVECTV 688
OY 718 ---DDLSVSEIVGKCGGEMRVASKRPRACRIDGEDVGFKKYDQO-MVYVQVPP 769
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 689 ETGEDGNAVVGLRVGRGFGAGYCSRPAKSVSDASADVEPTDGLVYADVPVP 744

```

RESULT 3
S45033
probable imbibition protein - wild cabbage
C:Species: Brassica oleracea (wild cabbage)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S45033
R:Fujikura, Y.; Karsen, C.K.
submitted to the EMBL Data Library, May 1994
A:Description: Cauliflower cDNA encoding a putative imbibition protein.
A:Reference number: S45033
A:Accession: S45033
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-765 <PU>
A:Cross-references: EMBL:X79350; NID:g488786; PID:g488787

Query Match 34.9%; Score 1473.5; DB 2; Length 765;
Best local similarity 37.6%; Pred. No. 1.9e-108;
Matches 297; Conservative 155; Mismatches 253; Indels 85; Gaps 18;

```

OY 21 MSSPFAIDGSDFTVNGHSLSDVPEINIVASPSPTYSIDKSPVSGCFGFDASEPDSRHV 80
      | : : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 3 ITNISVQNDNLVVOGKITLTKIPDNILIPVAGAGSDS-----GAFIGATFQKQSLHV 57
OY 81 VSLGKLKDTRFMSIFRFKVMWMTTHWGRNGDLESEIQIYLE-----SDSGRYVFL 134
      | | : : | | | | | : | : | : | : | : | : | : | : | : | : |
Db 58 FPLIGVLEGLREKMCFFRKLMWMTQRMGASGKDIPLTEQFLMSKDEVDNGDAPVYTYF 117
OY 135 LPIVEGPFRTSIOPGDDDEVDCVESGSSKVVVDASFRLMUTLHAGDDPFLVLEAKIVR 194
      | : : | | | : | : : | : | : | : | : | : | : | : | : | : | :
Db 118 LPLLEOFRVAVLQGNODELQICIESGDKAVETEOGMNWNVYIAGINPFTITQAVKAVE 177
OY 195 THLGFRLLLEKTRPGIIVDKFGMCTWDARFLVIVNPOGVLEGVNHVLDGCGRPLVILIDG 254
      | | : : | : | : | : | | | : | : | : | : | : | : | : | : | : |
Db 178 ROMQGTNHNREKKTVPFVDFWFGMCTWDARFLVIVNPOGVLEGVNHVLDGCGRPLVILIDG 237
OY 255 WQSIGH---DSDPTIKGAMQTYAGDOMPCRLKLFQENYKFRDYVNPKATGRPAQOKGMA 311
      | | : : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 238 WQOILEKKEKSNCLVQ-----GAOFATRLVGLIEEMAFQKN-DPKDT---QVSGLK 285
OY 312 AFIDELGKFEKTVENHYVNHALGCGWGLRPQVGLP--EARVITQPLSPGLQMTMEDLAV 369
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 286 SVYDNMAK-QHNHVKQYUAMHIALAGYWGKVPKSAAGMEHNDLSALATYPQVSGVLTGNQPDIVM 344

```



```

      8 SVYSDSLVYLGHVHLGVPENLVLPATSGNALID-----GAFIGVTSDDGTGSHRVFSLG 61
      85 KLADIFEMSIFFRRKVMWTHMVGRCNGLDLESEFQIVILEKS-----DSGRPYELLP 136
      62 KLELDLRFMCYFRKLMWMMQRMGTNGKEIPCEQFLIVEANOSDGLGRQSSSYVFLP 121
      137 IVEGPRFTSIQPGDDDFVDVCBSSSKYVDASFRLMVLHAGDDPFLVKEAMKIV-- 193
      122 ILEGDFRAVLQGNANELELICESGDPVTVDQFEGSHLVFVAAGSDPEVDITKAVKYMFL 181
      194 RFLHGF-----FRLLEKTPPGIVDKFGMCTDAFYLYVHPGVIEGVNHLVD----- 241
      182 KSQLKTSICLPNFFRM-----PDLNMFQCTWDADFTYNTADVQGLESNCDLTKPAL 235
      242 -----GGCPGLVLIDDGMOISIGHSDPITKEGNMQVAGQMCRCRLFOENKFRD 294
      236 ILCSLKAGGVTPEKFFVLIIDGMOVSGMDENSVEFNADN-----AANFANRLTHIKENHKFQ- 290
      295 YNPKATGPRAQCKMKA-----FIDELKGEFTVEHYVVMHALCGYWGGLRPQVP 345
      291 -----KDGKEGHRVDDPSSLFGHVLTDIKSN--NSLKYYVVMHALITGYMGVAKPGVS 340
      346 GLP--BARVIOPLSPGLQMTMEDLAVDKIVLHKVGIYPERKAEEMYEGLHAHLEKVGID 403
      341 GMEHYSKVAYPVSSPGVWSSSENGCLESITTKGGLVNPKEKVSFYNDLHSLASYAGVD 400
      404 GVKIDVTHLEMLCEDYEGGRVADLAKAYYKAMTKSINKHFKNGVISMHECHNFMFLGTE 463
      401 GYKVDQVONILETIGAGHGRVAKLAKKYHQALEISIRNFPDNGITISMSINTGLY-SAK 459
      464 AISLGRVGDFFWCTDPSGDPNGTFTWLQCGHMVHCAN--DSLWGNFTHPDMWFOSTHP 520
      460 KTVAVIASDDFMRDPA-----HTIHISVAYVNTLPLGEMQDMDMFHSLHP 508
      521 CAAFHAAARISISGPIYSDSVGKHNFDLKKIVLPDGLSKSEYALPFRDCLFEDPLH 560
      509 MAYEHAARAVGCAIYVSDRPGQHDNFNLKRLVLRDGSLLAKRLGPRPREDLY----- 563
      581 NGETMLKIMNLKFTGVIGAFNCGCGGMCETRRNOCFSQSKRYTSKTPKDIEMHSGE 640
      564 -----LPRD----- 567
      641 NPISIEGVKTFALYLYOAKKLLIKSPSODIALDPFEFELLIVSPYTKLIQTSLHRAPI 700
      568 -----TSLPYTIMPREVEVFVYVPEKPSDGS-KEAPV 599
      701 GLVNMILTSAGAIQSVYDDDLSS--VEIGVKGCGEMKRVAS-KKPRACRIDGSDVGFK 755
      600 GLMEMFNSSGAIYSLKRYDDEGTKEFVVRMKLRGSLGVGVSSVARRPSVTYVSDSDEYR 657

RESULT 6
T09530
probable seed inhibition protein - chickpea (fragment)
C:Species: Cicer arietinum (chickpea, garbanzo)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T09530
R:Cervantes, E.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z16718
A:Accession: T09530
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-357 <CER>
A:Cross-references: EMBL:X95875; NID:g1212811
A:Experimental source: germinating seed
C:Genetics:
A:Gene: sip

Query Match 11.6%; Score 490.5; DB 2: Length 357;
Best Local Similarity 32.5%; Pred. No. 6, 2e-31;
Matches 127; Conservative 55; Mismatches 154; Indels 55; Gaps 11;

```

```

      187 KEAMKIVRTHLGTFRLLEKTPPGIVDKFGMCTWDAFYLYVHPQVIEGVNHLVDGGCP 246
      3 QQAVKVAEKMHQFHHREKRVPSFLDMFGMCTWDAFYLYVHPQVIEGVNHLVDGGCP 62
      247 GLVILDDGMOISIGHSDPITKEGNMQVAGQMCRCRLK-----FQENYK--FRDIYNK 299
      63 RFLIM-----TWGNRLKLVKQSKSGCVGEGADQLGLDWMIRDA 101
      300 ATGPRAQOK-----GMAKAFIDELKGEFTVEHYVVMHALCGYWGGLRPQVPGLPE--AR 351
      102 NSKRNQNDQEQIPLGLKHLVDGK--KHNVKDYVVMHALAGYGVKPRPAATGEHNDAL 160
      352 VIQPVLSPLQMTMEDLAVDKIVLHKVGIYPERKAEEMYEGLHAHLEKVGIDGVKIDVH 411
      161 ATSAVTRSS--TWNQPDIVMSLAVHGLRLVHPRGFTSTNE--THAVLASGVDGKVDVQT 217
      412 LLEMLCEDYGGRVADLAKAYYKAMTKSINKHFKNGVISMHECHNFMFLGTEAISGRVG 471
      218 LRPPLVLDIVVESRLHAAILMRLLPLVTLNNGCIACMCHNTDGLY--SAKOTAIVRAS 276
      472 DDFWCTDPSGDPNGTFTWLQCGHMVH--CANDSLWGNFTHPDMWFOSTHPCAAFHAS 528
      277 DDFYPRDPA-----HTIHISVAYVNTLPLGEMQDMDMFHSLHRAEYHAAA 325
      529 RAISGPIYVSDSVGKHNFDLKKVLPDGS 559
      326 RAIGGCGQFMSVISQATITLFLRSVLVADGS 356

RESULT 7
T01717
hypothetical protein A_I6002N01.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
R:Schneel, P.; Magg, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana I6002N01.
A:Reference number: Z14407
A:Accession: T01717
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191156
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:Introns: 5/2; 25/3; 103/3; 126/1; 174/2; 240/1; 284/1; 295/1; 353/2; 349/2
A:Note: A_I6002N01.5

Query Match 11.2%; Score 471.5; DB 2: Length 371;
Best Local Similarity 26.9%; Pred. No. 2, 1e-29;
Matches 131; Conservative 71; Mismatches 112; Indels 173; Gaps 18;

```


Db 216 NSLKYVYVWHAITGYMGVKSPSVSGI-----MSNENCG----- 249
Oy 382 PEKAEWEYEGJLHALEKVGIDGVKIDVYHLEMLCEDYGGVADAKAYKAMTKSINKH 441
Db 230 -----LESTTKNGL-----GGGVKLAKKTKHQLLEMSISIN 279
Oy 442 FKNGCVIASMEHCNDFMELGTETASISLGRYDGFMCCTDPSGDNFTWLOGCHMVCAN-- 499
Db 280 FRANGKHNDLOYC-----SHSOKRQLQFMRHDPAS-----HTIHASVA 317
Oy 500 -DSLWKGNFIFHDWMFQST----HPCAFAHSAIRAISGPIYVDSVSGKNAFDLLKLV 554
Db 318 YNTLFLGEMFQPDMDISSSWRMCHLC-----QVYISDKPQOHDFNLLRKL 364
Oy 555 LPPGSIL 561
Db 365 LODGSIL 371

RESULT 8
D90496

hypotheetical protein SSO3127 (imported) - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: D90496

R:Sheng, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aweez, M.J.; Chan

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F

arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90496

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1 649 <KUR>

A:Cross-references: GB:AE006641; NID:q13816550; PIDN:AAK43227.1; GSPDB:GN00155

C:Genetics:

A:Gene: SSO3127

Query Match 8.6%; Score 365; DB 2; Length 649;
Best Local Similarity 24.0%; Pred. No. 1.4e-20;
Matches 156; Conservative 101; Mismatches 218; Indels 174; Gaps 32;

Oy 28 DGSDFTVNHSF--LSDVPENI-----VASPSPT-----SIDKSP 61
Db 37 DCKIYSIEBSKSVKISKRPITLKLNAKQDLSLTQOITYEVYKAFYVNOALIDMR 96
Oy 62 VSVGCFVGFDASEPDSR---HVSIGKLDKIDRFMSIFRFKVMWTTHWGRNGDLESETO 118
Db 97 V-----TEPKELKYDIKSLDHEKYIRTPCMLTPVF-----NNIPDYTV 136
Oy 119 IYILEKSDSGRYVFLPIVEGPRFTSIORGGDDFVDCVEGSSKVVVDSPRS-MLYH 177
Db 137 FALVSGNSYEAFETL---SNVYVATLEGD---SVALLYGFN--DEIKRSYFLSIG 186
Oy 178 AGDDPFAVLEAMKIVRTHLGTFRLLEEK--TPGVIDKRGMCMTWDAFYLT--VHPGVIE 234
Db 187 TSDNRYKALEMNIINASKTEFTFKLRKEKFPDKVYNGKMSMANF--LTRDLNENLTK 245
Oy 235 GVRHLVDGCGPGLVLIDGWSIGHSDPIYTKEGNQTUVAEQMPCRLKF--QENYK 292
Db 246 VYKGIIEIRLRLNWYLIIDGMQDQND-----RAIRSLPNPKKF 285
Oy 293 RQYVNPRAKPRAGOKGMAFIDELKGEFKYEHVYVMAHLCGYNGGLRPQVGLPEARV 352
Db 286 -----PNGFNVTYRAIKS--LGKYYGGLMAINAHWGMS----- 318
Oy 353 IOPVLSPIQMTMEDLAVDKIVLHKVGLVP---EKAEMEYEGJLHALEKVGIDGVKID 408
Db 319 -----QELMKSLNVCYFTNPLNSVPSPNLEDAIGFYKAPDGNLRL-DPDLVKYD 368
Oy 409 ---VTHLEMLCEDYGG-RVDLAKAYK--AMTKSINKHFKGNGVIASM-----EHCNDM 458
Db 409 ---VTHLEMLCEDYGG-RVDLAKAYK--AMTKSINKHFKGNGVIASM-----EHCNDM 458

Db 369 NQWYIHAI-----YDSEPIGLASHNIQIALQYSV-----GKDVINCMSNPENYCN-YF 416
Oy 459 ELGTASISLGRYDGFMCCTDPSGDNFTWLOG--CHWYHCANDSLAMGNFTHPMDMO 516
Db 417 YSNWRNSTIDY-----PEWKDGYLHMFNAYNSLTSHTYVPTDPM 461
Oy 517 STHPCAAFHSAIRAISGPIYVDSVSGK----NFDLLKLVLPGLSILRSEYVALPTPD 572
Db 462 SYDPAKXHLVARVSGGYITD---RHPERTNELLRMALPLPGVIRVDEPALITED 518
Oy 573 CLFEDPLNGETMLKIWNLFKTYGIGAFCNCGGCMCRETRNOCPSOY 621
Db 519 LFKDPLRE-RVLLKTKGKVKGYNAIAEFNLNSGEVEEYNNEDYXX 566

RESULT 9

S76481
hypotheetical protein - Synechocystis sp. (strain PCC 6803)

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S76481

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima,

O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocys

s.

A:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S76481

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-538 <KAN>

A:Cross-references: EMBL:D90915; GB:AB001339; NID:q1653604; PIDN:BA18610.1; PID:q101

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 2.9%; Score 121.5; DB 2; Length 538;
Best Local Similarity 21.5%; Pred. No. 0.22;
Matches 109; Conservative 79; Mismatches 185; Indels 135; Gaps 29;

Oy 21 MSSPRAIDSDFTVNGHSLSDVPENIVASPSPIYSIDKSPVSGCFVFDASEPD---- 76
Db 19 IASPSVYKGA-----ETTFISIMPLQOFDISVSLTDFATGTIDDPKRY 64
Oy 77 SRHV--VSIGKLDKI-----RPSIFRFKVMWTTHWGRNGDLESETOIVILEKSDSGR 129
Db 65 TQHLNPELEKLRGLNHSFKNSVAFREFNTTF-----GKEIAQQLSYITAAPIDOSQ 119
Oy 130 PYVFLPLPIVEGPRFTSIQ--PGDDFVDCVEGSSSKVY---DASFRSLYLHAGDDPFA 185
Db 120 PE-----LEGATVTAQAQNPDGFKIIDIVINAVGSDLVNLMD--TEKNTI-----DOADT 166
Oy 186 VKEAKIVYTHLGTFRLEEKTPPGI---VDFGKCTYDARYLV-HQG-VIEGVRIL 239
Db 167 LVQATDRIFTWLGKOEIPSNPVPMLKPLALAEPOGQKKTQNTLTIIPRNGPVNVFVL 226
Oy 240 VGGCGPFG-LVLIDG---WQSIGHSDPIYTKEGNQTUVAEQMPCRLKF--QENYK 293
Db 227 PGQNSPRLVYVIAAGLNSNFOAFYTIADHLASYG--TIAGIDPESAAKQDSLOGL 284
Oy 294 D-YVNPRAKPRAGOKGMAFIDEL-----KGEFKYEHVYVMAHLCGY----- 336
Db 285 DAFPDPMALQO--PRDVTLVLDTLAQAKATDPAMQKFDINNVGILGHSGLGYATAISG 342
Oy 337 -----WG-----GLRQVYGLPEARV-----IQVLS 358
Db 343 GATLEMPBELLQOCEOLNKNQINLPALLMOCOGVGSAPASNLRSRIRKAVLAINPVN 402
Oy 359 P-----GLQMTMEDLAVDKIVL--HKVGLVPEKAE---EMYEGJLHALEKVGIDGVKID 407
Db 403 PIFGPRG---WKNLAVPLLIYAGSKDIFARPVPDQIIFSLIEGVNKKLLLYO--NGTHL 457
Oy 408 DVTHLEMLCED-YGGRVDLAKAYKAM 434
Db 408 DVTHLEMLCED-YGGRVDLAKAYKAM 434

A:Cross-references: SGD:S0002827; MIPS:YDR419w
A:Map position: 4R

Query Match 2.7%; Score 113.5; DB 2; Length 632;

Best Local Similarity 21.2%; Pred. No. 1.2; Mismatches 197; Indels 169; Gaps 23;

Matches 113; Conservative 55; Mismatches 197; Indels 169; Gaps 23;

```

QY 277 EQMPCKLLK-----FOENYKFRDYVNPKATGPACGKAKAFIDELKGEFTV 324
DB 39 EQMRGLSEDEPVVVCVQNMSTIAVSAARKYGSMDTIQELAKCSNLIPIHTAVFKKG 98
QY 325 EHYVYVHVALCGYWGGLRPQGLPEARVIOVLSQGLQMTMEDLAVDKIVLHKVGLVP-- 382
DB 99 EDFWQYHDCGSGW-----VODPAKQISVED-----HKVSLPEYR 132
QY 383 --PEKAEEMYEGELHMLEKVGIDGVKIDYIHL--LEMCEDYCGRVDLAKAYKMTKS 437
DB 133 RSRRAALKIFKASCDLVERASIDEVFLDGRICFNMMLMDNEVELTGDLK--LKDALSN 189
QY 438 INKHFKNCGVISMHECHNDFMELGTEAI-SLGRVGDDEFWCTPSPGDPNGTFW----- 488
DB 190 IREAFIGWY-----DINSHLPLPEKIKSLKEGDF--NPEGRLTTDMDDVILAAG 241
QY 489 -----LOGCHMVHCANDSLMNGNFIHPDW-----DMFOSTHPCAERH 525
DB 242 SOVCKGIRDINDILGYTTSCLSTSKNVCKLASNYKKPDQITVKNDCLLDFLDGKFE 301
QY 526 AASRAISGPIYVSDSVGKHNFDLLKLLP-----DGSILR 562
DB 302 ITSPWTLGVGL-----GKELIDVLD--LPHENSIKHIRETPWPNAGLKEFLDKVKQ 352
QY 563 SEVYALPFRDCLEDFPLHNGETMLKIMNLN-----FTVGIAFNOCGGG--- 607
DB 353 SDY-----DSTSNIDPLKADLAELKFLSRGYGLPLSSRPVVKSMNKNLKGSCNS 408
QY 608 -----WCRE-TRRNOCFSSQ-YSKRVTSKT-----NPKDIEHSGENPISIEGVKT 650
DB 409 IYDCISMLEVFCAELTSRIQDLEQENKIVIPRTVASISLKTSEYEVRRKSGPAAVKGIN- 467
QY 651 FALVLYQAKLLI--SKPSQDLDIALDPPEFELITYSPYTKLIQTSYLHAPICL 702
DB 468 -----FOSHELKLVGIKVYTDLDI-----KGKKSTYPLTKLSMTITNPDIDL 511

```

RESULT 13

164118

1.4-alpha-glucan branching enzyme (EC 2.4.1.18) - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: 164118

R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gockyne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: 164118

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-730 <TIGR>

A:Cross-references: GB:U32815; GB:L42023; NID:g1574818; PIDN:ANC23004.1; PID:g1574820; T

C:Superfamily: 1.4-alpha-glucan branching enzyme

C:Keywords: glycogen/starch biosynthesis; glycosyltransferase; hexosyltransferase

Query Match 2.5%; Score 105; DB 1; Length 730;

Best Local Similarity 17.5%; Pred. No. 7.3; Mismatches 203; Indels 272; Gaps 29;

Matches 117; Conservative 75; Mismatches 203; Indels 272; Gaps 29;

```

QY 68 VGDASEPDSRHVSTIGKLDIRFMSIFRPKWMTTHWGRNG--DDESETQIYLK 124
DB 138 VNRFLMAPNARRSVIGDF-----NYWDCRRHPRFHSKSGVWELFLPK 181

```

```

QY 125 SDGRRYVELLPIVBS-----PRTSIQPDDEFDVDCVESGSSKVVADAFSRMLYLH 177
DB 182 ASLQGLYKRELLDCHGNLKLKADPFASFSSQLRPDPAQSVASLPVNVEMEARKKA-----N 237
QY 178 ACDDEPALVKEAMKYRTHLGTFRLLLEETPPGIYDKFQWCTWDAFYLTVPQYIEGR 237
DB 238 QGNOP-----ISYEVHLSMR-----RNLENMF-WLDDYQIADLILPYVKEMGPT 282
QY 238 HL-----VDGGRPELVILIDGQWOSIGHSDSDITKEGNNQVYAGQMCRLLKPE 288
DB 283 HLEPLSEFPDGSW-----GYOPLGLYS-PTSFGSPEAFR-----RLVK--- 323
QY 289 NYKFRDYVNPKATGPACGKAKAFIDELKGEFTVENHYVHVALCGYWGGLRPQGLP 348
DB 324 -----RAHAGINVLIDWVPGHFPDTH----- 346
QY 349 EARVIOVLSQGLQMTMEDLAVDKIVLHKVGLVPRPEKAEEMTEGLAH----- 396
DB 347 -----GL-----VAEDGTALY-----EHEDPREGYHODMNTLTYNGRNE 381
QY 397 -----LEKVGIDGVKIDYIHL-----EMCEDYCGRVDL-AKAYKA 433
DB 382 VKNFLSSNALYWERFGVGIHYDAVASMTYRDYSAEBEMIPNOTGRENLEALEFLKH 441
QY 434 MKRSINKHFKNCGVISMHECHNDFMELGTEAISLGRVGDDEFWCTPSPGDPNGTFWLOGCH 493
DB 442 TMMKI--HSEMGALSIARESTSPAGV--THPSENGSLGNF-----KMMMG-- 484
QY 494 MVHCANDSLMNGNFIHPDWDMFOSTHPCAFAHAAISAGGPITYSDSVGKHNFDLLKLL 553
DB 485 -----WMNDTL-----AYMKLDPIRYOHHNKKMIFGVHYOY 515
QY 554 VLPDGSILSEVYALPFRDCLEDFPLHN-----GFMKIMNLNKEFTYGIAF-- 601
DB 516 -----SENFVLPLS--HDEVHGYSLGKKRPQWFOFALVRAVYIGIMGYPRK 563
QY 602 -----NCQ-----GGGMCRETRR-----NOCPSQYSKRVTSKT 629
DB 564 KLFPMGNEFAQGRNMYEESLDWFLDENIGGSMHNGVGLVKVLDNQIYOKNNRPLFELDN 623
QY 630 NPKDIEW 636
DB 624 SPEGFDW 630

```

RESULT 14

T19997

hypothetical protein C47D12.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T19997

R:Gajadasy, S. submitted to the EMBL Data Library, March 1996

A:Reference number: Z19209

A:Accession: T19997

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-3944 <WIL>

A:Cross-references: EMBL:Z69902; PIDN:CA93765.1; GSPDB:GN00020; CESP:C47D12.1

A:Experimental source: clone C47D12

C:Genetics:

A:Gene: CESP:C47D12.1

A:Map position: 2 A:introns: 46/3; 308/3; 408/2; 449/3; 820/3; 946/3; 1069/1; 1634/3; 1737/2; 2009/2; 2

Query Match 2.4%; Score 103; DB 2; Length 3944;

Best Local Similarity 18.5%; Pred. No. 1.4e+02; Mismatches 136; Conservative 90; Mismatches 230; Indels 280; Gaps 34;

Matches 136; Conservative 90; Mismatches 230; Indels 280; Gaps 34;

```

QY 80 VVSIKLLKIRPSTIRKRVWMTTHWGRNGDLESETQIVILEKSDSGRPYVFLPIVE 139
DB 2931 VVSYGHVKITLQAMNIVR-----EIEESTDIRI----- 2957

```


GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:01:51 ; Search time 76 Seconds
(without alignments)
1374.585 Million cell updates/sec

Title: US-09-675-208-5
Perfect score: 4225
Sequence: 1 MAPSPKNGSNVSPFDGLND.....QVMPRPDSSGGISVIEYLF 784

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_101002.*
1: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgcdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgcdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgcdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgcdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4225	100.0	784	19 AAW53570	Cucumber raffinose
2	4225	100.0	784	20 AAY17417	Cucumber raffinose
3	2996.5	70.9	777	20 AAY32074	Mustard raffinose
4	2945	69.7	783	23 ABB93664	Herbicidally active
5	2860.5	67.7	783	20 AAY32073	Sugarbeet raffinose
6	2854.5	67.6	781	19 AAW57887	Soybean raffinose
7	2854.5	67.6	781	20 AAY30143	Amino acid sequenc
8	2854.5	67.6	781	20 AAB49400	Soybean raffinose
9	2852	67.5	758	21 AAY70978	Soybean raffinose
10	2834	67.1	780	22 AAB98659	Soybean protein: S

11	2767	65.5	799	19 AAW57886	Broad bean raffin
12	2767	65.5	799	20 AAY30142	Amino acid sequenc
13	2529	59.9	587	19 AAW57888	Japanese artichoke
14	2529	59.9	587	20 AAY30144	Amino acid sequenc
15	2364	56.0	572	20 AAY32075	Rapeseed raffinose
16	1853.5	43.9	807	23 ABB92806	Herbicidally active
17	1490	35.3	751	21 AAY70980	Wheat raffinose sy
18	1470	34.8	763	21 AAY70977	Rice raffinose syn
19	1432.5	33.9	756	21 AAY70975	Corn raffinose syn
20	1398.5	33.1	841	21 AAY70981	Wheat raffinose sy
21	1394	33.0	750	20 AAY17418	Soybean raffinose
22	1384	32.8	770	21 AAY70976	Rice raffinose syn
23	940	22.2	508	21 AAY70974	Corn raffinose syn
24	873	20.7	265	20 AAY32072	Soybean raffinose
25	788	18.7	271	19 AAW57889	Corn raffinose syn
26	788	18.7	271	20 AAY30145	Amino acid sequenc
27	389.5	9.2	229	21 AAG28628	Arabidopsis thalia
28	263.5	6.2	193	21 AAG28629	Arabidopsis thalia
29	181	4.3	30	19 AAW53567	Cucumber raffinose
30	181	4.3	30	20 AAY17422	Cucumber raffinose
31	177.5	4.2	120	21 AAB28630	Arabidopsis thalia
32	153	3.6	63	23 AAB25825	Human ORF4040 prot
33	111	2.6	1174	13 AAR25825	Novel toxin expres
34	109	2.6	1390	18 AAW21636	Grapevine leafroll
35	107.5	2.5	1584	20 AAY33727	Photobacterium lum
36	107	2.5	487	19 AAW64182	Lettuce resistance
37	107	2.5	487	23 AAW95497	Lettuce pest resis
38	107	2.5	488	19 AAW64159	Lettuce resistance
39	107	2.5	488	23 AAW95482	Lettuce pest resis
40	107	2.5	1174	11 AAR06257	B. thuringiensis to
41	107	2.5	1174	17 AAR89494	B. t. toxin 81A2.
42	107	2.5	1174	18 AAW09043	Bacillus thuringie
43	106.5	2.5	905	22 AAB20088	Rat CARD-6 protein
44	106	2.5	1080	22 AAM00803	Human bone marrow
45	106	2.5	1372	22 AAG14534	Novel human diagno

ALIGNMENTS

XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX	XX
----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA:

Query Match 100.0%; Score 4225; DB 19; Length 784;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSFNKGSNVYSPFDGLNDMSSPFAIDGSDFTVNGHSLSDVPENIVASPSPTYSIDKS 60
 DB 1 MAPSFNKGSNVYSPFDGLNDMSSPFAIDGSDFTVNGHSLSDVPENIVASPSPTYSIDKS 60
 QY 61 PVSVCGFVGFDASEPSRHHVSIKGLKDIRFMSIFRFKVMWTTTHWYGRNGDLESETQIV 120
 DB 61 PVSVCGFVGFDASEPSRHHVSIKGLKDIRFMSIFRFKVMWTTTHWYGRNGDLESETQIV 120
 QY 121 ILEKSDSGRPVYFLPIVEGPFRTSIQPGDDDFVDVCVESSSKVVDASFRSMLYLHAGD 180
 DB 121 ILEKSDSGRPVYFLPIVEGPFRTSIQPGDDDFVDVCVESSSKVVDASFRSMLYLHAGD 180
 QY 181 DPFAIVKEMAKIVRTHLGFTRLLSEKTPRGIVDKFGMCTWDAFYLTIVHQVIEGVRLHY 240
 DB 181 DPFAIVKEMAKIVRTHLGFTRLLSEKTPRGIVDKFGMCTWDAFYLTIVHQVIEGVRLHY 240
 QY 241 DGGCPRPGLVLIIDGWSIGHSDSPITKEGNOTVAGEOMPRLKFOENKFRDYVNPRA 300
 DB 241 DGGCPRPGLVLIIDGWSIGHSDSPITKEGNOTVAGEOMPRLKFOENKFRDYVNPRA 300
 QY 301 TGPRAQOKMKAFIDELKGEFKEVHEHYVWHALCGYWGGLRPQGLPEARVYIQVPLSPG 360
 DB 301 TGPRAQOKMKAFIDELKGEFKEVHEHYVWHALCGYWGGLRPQGLPEARVYIQVPLSPG 360
 QY 361 LQMTMEDLAVDKIVLHKVGLVPRPEKAEMYEGILHAHLEKVGIDGVKTDVYIHLEMLCEDY 420
 DB 361 LQMTMEDLAVDKIVLHKVGLVPRPEKAEMYEGILHAHLEKVGIDGVKTDVYIHLEMLCEDY 420
 QY 421 GGRVDLAKAYYKAMTKSINKHFRKNGVIVASMEHCNDFMFLGTEAISLGRVGDPMCTDPS 480
 DB 421 GGRVDLAKAYYKAMTKSINKHFRKNGVIVASMEHCNDFMFLGTEAISLGRVGDPMCTDPS 480
 QY 481 GDRNGTFWLOGCHMYVHCANDSLMGNFIHPDMDMFQSTHPCAFAHSAASRAISGGPIYVSD 540
 DB 481 GDRNGTFWLOGCHMYVHCANDSLMGNFIHPDMDMFQSTHPCAFAHSAASRAISGGPIYVSD 540
 QY 541 SVGKHNFDDLKTLVLPDGSILNSEYVALPTROCLFEDPLHNGETMLKIVNLKFTGVIGA 600
 DB 541 SVGKHNFDDLKTLVLPDGSILNSEYVALPTROCLFEDPLHNGETMLKIVNLKFTGVIGA 600
 QY 601 FNCGGGMCRETRRNQCFQYSKRVTSKTPNKDIEHSGENPISIGVTFALYIQAAK 660
 DB 601 FNCGGGMCRETRRNQCFQYSKRVTSKTPNKDIEHSGENPISIGVTFALYIQAAK 660
 QY 661 LILSKPSQDLIALDPFEELITVSPVTKLIQTSLHFAPIGLVNMNTSGAIOQVYDD 720
 DB 661 LILSKPSQDLIALDPFEELITVSPVTKLIQTSLHFAPIGLVNMNTSGAIOQVYDD 720
 QY 721 LSSVEIGVAGCEMRVFAKKPRACIDGEDVGFKIDQOMVYVYVPMPIIDSSSGTISYI 780
 DB 721 LSSVEIGVAGCEMRVFAKKPRACIDGEDVGFKIDQOMVYVYVPMPIIDSSSGTISYI 780
 QY 781 EYLF 784
 DB 781 EYLF 784

RESULT 2
 AAY17417
 ID AAY17417 standard: Protein; 784 AA.

XX AC AAY17417;
 XX 29-JUL-1999 (first entry)
 DE Cucumber raffinose synthase.
 KW Raffinose synthase; sucrose; galactinol.
 XX Cucumis sativus.
 OS JP1123080-A.
 PN 11-MAY-1999.
 PD 24-OCT-1997; 97JP-0292969.
 PF 24-OCT-1997; 97JP-0292969.
 PR 24-OCT-1997; 97JP-0292969.
 PA (AJIN) AJINOMOTO KK.
 DR WPI: 1999-340516/29.
 DR N-PSDB; AAX61238.
 PT New raffinose synthase gene - for production of raffinose from
 PS sucrose and galactinol
 XX Claim 2; Page 25-27; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents raffinose
 CC synthase from cucumber.

XX Sequence 784 AA:

Query Match 100.0%; Score 4225; DB 20; Length 784;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAPSFNKGSNVYSPFDGLNDMSSPFAIDGSDFTVNGHSLSDVPENIVASPSPTYSIDKS 60
 DB 1 MAPSFNKGSNVYSPFDGLNDMSSPFAIDGSDFTVNGHSLSDVPENIVASPSPTYSIDKS 60
 QY 61 PVSVCGFVGFDASEPSRHHVSIKGLKDIRFMSIFRFKVMWTTTHWYGRNGDLESETQIV 120
 DB 61 PVSVCGFVGFDASEPSRHHVSIKGLKDIRFMSIFRFKVMWTTTHWYGRNGDLESETQIV 120
 QY 121 ILEKSDSGRPVYFLPIVEGPFRTSIQPGDDDFVDVCVESSSKVVDASFRSMLYLHAGD 180
 DB 121 ILEKSDSGRPVYFLPIVEGPFRTSIQPGDDDFVDVCVESSSKVVDASFRSMLYLHAGD 180
 QY 181 DPFAIVKEMAKIVRTHLGFTRLLSEKTPRGIVDKFGMCTWDAFYLTIVHQVIEGVRLHY 240
 DB 181 DPFAIVKEMAKIVRTHLGFTRLLSEKTPRGIVDKFGMCTWDAFYLTIVHQVIEGVRLHY 240
 QY 241 DGGCPRPGLVLIIDGWSIGHSDSPITKEGNOTVAGEOMPRLKFOENKFRDYVNPRA 300
 DB 241 DGGCPRPGLVLIIDGWSIGHSDSPITKEGNOTVAGEOMPRLKFOENKFRDYVNPRA 300
 QY 301 TGPRAQOKMKAFIDELKGEFKEVHEHYVWHALCGYWGGLRPQGLPEARVYIQVPLSPG 360
 DB 301 TGPRAQOKMKAFIDELKGEFKEVHEHYVWHALCGYWGGLRPQGLPEARVYIQVPLSPG 360
 QY 361 LQMTMEDLAVDKIVLHKVGLVPRPEKAEMYEGILHAHLEKVGIDGVKTDVYIHLEMLCEDY 420
 DB 361 LQMTMEDLAVDKIVLHKVGLVPRPEKAEMYEGILHAHLEKVGIDGVKTDVYIHLEMLCEDY 420
 QY 421 GGRVDLAKAYYKAMTKSINKHFRKNGVIVASMEHCNDFMFLGTEAISLGRVGDPMCTDPS 480
 DB 421 GGRVDLAKAYYKAMTKSINKHFRKNGVIVASMEHCNDFMFLGTEAISLGRVGDPMCTDPS 480

AB93664
ID AB93664 standard: Protein; 783 AA.
XX AC AB93664;
XX DT 31-MAY-2002 (first entry)
XX DE Herbicidally active polypeptide SEQ ID NO 2875.
XX KW Herbicidal; plant; agriculture; herbicide.
XX OS Arabidopsis thaliana.
XX PN WO200210210-A2.
XX PD 07-FEB-2002.
XX PF 28-AUG-2001; 2001MO-EP09892.
XX PR 28-AUG-2001; 2001MO-EP09892.
XX PA (FARB) BAYER AG.
XX PI Tietjen K, Weidner M;
XX DR WPI: 2002-269010/31.
XX PT Identifying plant target proteins for herbicidally active compounds,
XX PR comprising aligning and comparing nucleic acid or amino acid sequences
XX PT from plant with nucleic acid or amino acid sequences from non-plant
XX XX organisms -
PS Claim 5; SEQ ID NO 2875; 261pp + Sequence Listing; English.
XX XX
CC The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for hercidially active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX XX
SQ Sequence 783 AA:
Query Match 69.7%; Score 2345; DB 23; Length 783;
Best Local Similarity 69.2%; Pred. No. 1,9e+269;
Matches 544; Conservative 99; Mismatches 117; Indels 26; Gaps
OY 14 SFDGLN--DMSSPFAIDGSDFTVNGHSLDSVPENIVASPSPYTSIDKS---PVSVGR 67
: : : : : : : : : : : : : : : : : : :
DB 9 SDSGANGVDVFTEKFRLEDSTLLANGQVLTDPVNVTLTSSPYL-VDKDGVPLDVASGSF 67
OY 68 VGFDA-SPEDSHHVYSIGLKLDIRFSITRFKYWWTTHVGRGGELESETQVILEK-- 124
: : : : : : : : : : : : : : : :
DB 68 IGFNLDEGRSHHAVSIGLKLNIFRMSIFRFYVMWTTHWGSGNRGIENETQIIIDQSG 127
OY 125 -----SDSGRPVFLRPVEGPFRISIQPGDDDFVDCVESGSKRVNDVASPRMLYLHA 178
DB 128 SDSGSGSGSGRRYVLLPLLEGSSFRSSFOGEDDDVAVCVEGSESTVTGSEFRQIYYHA 187
OY 179 GDPPALVKEAMKIYRTHLGTFRLLEEKTTPGIVDKFGMCTMDAFLYLTHPOGIVEGNH 238
DB 188 GDPPKLVYDAKWAVIRVHNHTFKLLEEKSPRGIVJVKFGMCTMDAFLYLVNPGVHNKGVC 247
OY 239 LVDGCGCPGLVLIDGWOSIGHDSPITKEGMNQYVAGQMPCRLKFDQENTKFRDYVNP 298
DB 248 LVDGCGCPGLVLIDGWOSIGHDSPIDGEKNITVAGQMPCRLKFEENHKFYVSP 307
OY 299 KATGRRAQGKGAKAFDELKGEGFTEHYVNHAIQGYWGSLRPQVPGLEARVLIQPVLS 358
: : : :
DB 308 K-----DQNDVGKAIVRDLDKDEFSTVDYLYVNHAIQGYWGSLRPKPAPLPSTILPELS 363

OY		359	PGDMFMEDLAVNDKIVLVAKHGLVAPPEKAEEKYEGEIAHLLEKVGDGKIDVTIILLEMCE	418		
Db		364	PGLTKLTMDLVDVKIIETGEFASPDLAKEFYBGLSHLONAGIDGVKKVVIHILEMCO	423		
OY		419	DYGGHVDLAKAYVRAMTKRSIKRHKFGNGVIASMECHNDFELGTETAISLRVGDDFWCTD	478		
Db		424	KYGGVVDLAKAYFEVALITSVVKHKFNMGVLIASMCHCNDFELGTETAISLRVGDDFWCTD	483		
OY		479	PSGDPNGTFWLOCGHMVHCANDSIMGNFIHPMDMFQSTHPCAFHAASRAISGPITY	538		
Db		484	PSGDPNGTFWFMOGHMVHCAYSNSLMGNFIOPDWDMFOSTHPCAFFHAASRAISGPITY	543		
OY		539	SDSGKHNFDDLKTLYLPDGSILRSEYYALPTRCLPEDPLHNGETMLKIWNINKFPGVI	598		
Db		544	SDCVGRKHFDLTKRLVYLNGSILICEYYALTPTRLLEDPLHDOKTMLKIWNINKFTGYI	603		
OY		599	GAFNOCGGCMCRERLRNDCFSQYSKRRTSKTPNPDIEMHSGENPISIEGYKTEALXYQA	658		
Db		604	GAFNOCGGCMCRERLRNDCFSQCYNTLATTSXPDKVEMSGSSPISIANVEEFALFLSQS	663		
OY		659	KKLILSRKSODLIALDPEFELLITVSPLYKLLOTSLHFAPIGLVNNLNTSGAISOVDVD	718		
Db		664	KKLLLSGINDDLETLEBPFKELLITVSPVYTIEGNSVPRAFPIGLVNNLNTSGAIRSLVN	723		
OY		719	DOLSEVEIGVGCCEMRFAEKFRACRIDEDYGFFYDDODOMVVQVPPRISSSGGIS	778		
Db		724	DE--SVENGVGAGEFRRYAKKRVSCLDIDGVEYEFOT-EDSNVMQVPM---SGPDGLS	777		
OY		779	VIEYLE 784 ::			
Db		778	SIOYLF 783			
RESULT 5						
AA132073						
ID	AA132073	standard; Protein: 783 AA.				
XX	AA132073;					
XX	AA132073;					
XX	17-JAN-2000	(first entry)				
De	Sugarbeet raffinose synthase.					
XX	Raffinose synthase; sugarbeet; transgenic plant.					
XX	Beta vulgaris.					
OS						
XX	Key	Location/Qualifiers				
FH	Misc-difference 227	/note= "encoded by CCW"				
FT	Misc-difference 348	/note= "encoded by CCR"				
FT						
XX	EP953643-A2.					
PN	03-NOV-1999.					
XX	03-NOV-1999.					
XX	27-APR-1999;	99EP-0107430.				
XX	30-APR-1998;	98JP-0120550.				
PR	30-APR-1998;	98JP-0120551.				
PR	04-DEC-1998;	98JP-0345590.				
PR	10-DEC-1998;	98JP-0351246.				
XX	(SUMO) SUMITOMO CHEM CO LTD.					
PA	Watanabe E, Oeda K;					
PI	WPI: 1999-593144/51.					
XX	N-PSDB: AAZ20208.					
DR	New sense and antisense genes, useful for altering the level of					
XX						

PT raffinose in food plants -
 XX
 PS Claim 25: Page 22-24; 55pp; English.
 XX
 CC This sequence represents sugarteet raffinose synthase, a protein
 CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
 CC hydroxy group attached to the carbon atom at the 6-position of the
 CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
 CC (see AA20208) encoding the enzyme was isolated from sugarbeet cv.
 CC hamling leaf cDNA by PCR. Probes or primers generated from plant
 CC raffinose synthase genes (see AA20207-10) may be used to obtain
 CC other raffinose synthase genes by labeled detection or amplification
 CC (claimed). These genes may be used to control the levels of
 CC raffinose produced in plants. Antisense genes can be used to knock
 CC out existing gene activity, and sense genes to increase the level
 CC of gene activity. The resulting transgenic plants may be used as a
 CC food source to alter the growing conditions for gut enterobacteria,
 CC providing general health advantages.
 XX
 SQ Sequence 783 AA:
 Query Match 67.7%: Score 2860.5; DB 20; Length 783;
 Best Local Similarity 66.6%: Pred. No. 2e-261;
 Matches 530; Conservative 109; Mismatches 132; Indels 25; Gaps 12:
 OY 1 MAPSFKNKGS---NVVSEFDGLDMSSP-FAIDGSDFTVNGHSFLSDVPENIVASPSPYT 55
 DB 1 MAPSFKNKSKTCDEVAHNDHCN--TCPIISLESNFVNGHILISQVPSNI-----T 51
 OY 56 SLDKSVSVGCGYGFDPASBPDSRHVVSIGKLDIRFMSIFREKVMWTTTHVWGANGDLES 115
 DB 52 AISKMGFD-GLFVGFPAPKARHVVSVGLKIPMSIFREFVWMTTHTGSGNGRDLH 110
 OY 116 EPGIVLEKSDS--GRPVFLPIVEGPFRTSTQPGD-DDFVNVCESSGSKVVDASPRS 172
 DB 111 EPGIILLDSDSEGLRPYVILPLIEGPRASIQPGSDVDYDVCESGTVKVGGSFRA 170
 OY 173 MLYLAHADDPFALVPEAMKIVRTHLGTFRLLEKTPPGIIVDKFQMTWDAFYLTVHPGV 232
 DB 171 VYIRAGPPPFKLKIDTKMEVQAHLGTFKLLDKTPPGIIVDKFQMTWDAFYLVKPEYGV 230
 OY 233 IGVNRILVGGCPGPGVLLDDGWSIGHSDDPTT-KEGAKQVYAGQDMPCRLKFDENK 231
 DB 231 WEGVNLVNGVPPGLVLLDDGWSIGCHDDPTTDEGINRISAGQMPRLKYENEFK 230
 OY 292 FRDYVPRKATGPR-AGOKGMAFIDELKGEFTEVHEVYVWALCGWGLRPOVGLPEA 350
 DB 291 FRDYKSPNIMGHEDHNMGRARVRLKEEFKEVHEVYVWALCGWGLRPOVGLPEA 350
 OY 351 RVIQPVLSFGLQMTMEDLAVDKIVLHKVGLVPRPEKAEMEYEGLAHLKVGIDGVKIDVY 410
 DB 351 QVVTPLKSLGLEMTMEDLAVDKIVNNGIGLVOPDKAOELTEGSHLENGIDGVKVDVY 410
 OY 411 HLEEMCEGYGGVNDLAKAYYKAMTKSINKHFGNGVIVASMEHCNPMFLGTATISLGRV 470
 DB 411 HLEEMAEEDYGGVNEALAKTYKAITESVVRHFGNIVIVASMEHCNPMFLGTETIGRAV 470
 OY 471 GDFNCTDSDGPNCTFWMLOGCHMVHCANDSLMGNFIRHDMDFOSTHPCAFHAASRA 530
 DB 471 GDFNCTDSDGPNCTFWMLOGCHMVHCANDSLMGNFIRHDMDFOSTHPCAFHAASRA 530
 OY 531 ISGGPIYVSDSVGKHNFDLKLKLVLDGSGILRSEYVALPTROCLFEDPLHNGETMLKWN 550
 DB 531 ISGGPIYVSDSVGKHNFDLKLKLVLDGSGILRSEYVALPTROCLFEDPLHNGETMLKWN 550
 OY 591 LNKFGVITAFNNGGCGMRETRNOCFQSKRVTSKPNKDIEMHSGENPISIGVKT 650
 DB 591 LNKFGVITAFNNGGCGMRETRNOCFQSKRVTSKPNKDIEMHSGENPISIGVKT 650
 OY 651 FALYLOAKKLILSKPSODLIALDPEFELIVSPVTKLIQTSLHFAPIGLVNLMLNTSG 710
 DB 651 FALYLOAKKLILSKPSODLIALDPEFELIVSPVTKLIQTSLHFAPIGLVNLMLNTSG 710

OY 711 AIGSDVYDDDLSS--VEIGVKGCGENKRVASKKPRACRIDGEDGVGFKYDQDQWVVVQVPW 768
 DB 711 AVKSLDISEDNEDKMYOVGKAGKEMVYVSSKPKACRVNGEDMEERY-EESNIXQVTVW 769
 OY 769 PIDSSSGISVIEYLF 784
 DB 770 --NHNSGGEFTVEYLF 783
 RESULT 6
 AAW57887
 ID AAW57887 standard; Protein; 781 AA.
 AC AAW57887;
 DT 23-SEP-1998 (first entry)
 DE Soybean raffinose synthetase.
 KW Raffinose synthetase; metabolism modification; food additive;
 KM gastrointestinal flora; soybean.
 OS Glycine max.
 PN EP849359-A2.
 PD 24-JUN-1998.
 PF 18-DEC-1997; 97EP-0122417.
 PR 18-DEC-1996; 96JP-0338673.
 PA (SUMO) SUMITOMO CHEM CO LTD.
 PI Oeda K, Wantanabe E;
 DR WPI: 1998-324670/29.
 DR N-PSDB: AAW40801.
 PT New nucleic acid molecule encoding plant raffinose synthetase -
 PT capable of producing raffinose, used as food additives with
 PT beneficial effects on gastrointestinal flora
 XX
 XX Claim 1: Page 31-34; 44pp; English.
 CC
 CC This sequence represents the soybean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.
 XX
 SQ Sequence 781 AA:
 Query Match 67.6%: Score 2854.5; DB 19; Length 781;
 Best Local Similarity 65.9%: Pred. No. 7.2e-261;
 Matches 525; Conservative 123; Mismatches 120; Indels 29; Gaps 11:
 OY 1 MAPSFKNKGSNVVSPD--GLNDMSSEFAI--DGSDFTVNGHSFLSDVPENIVASPSPYTS 56
 DB 1 MAPSI-----SKVELNSFGLVNGNLPISLTIEGNSNFLANGHPFLTEVPEIITYPSP--- 53
 OY 57 ID-KSPVS-----VCGPFGDASBPDSRHVVSIGKLDIRFMSIFREKVMWTTTHVWGRN 109
 DB 54 IDAKSSKNNEDDVYCCFVGFADEPRSRHVASLGRKIGKFMISIRFKVMWTTTHVWGSN 113
 OY 110 GGLSESTOIVILEKSDS--GRPVFLPIVEGPFRTSTQPGD-DDFVNVCESSGSKVVD 168
 DB 114 GHLEHETQWMLDKNDQOLGRPVVLLPILOASFRASLOPGLDDYDVYDVCESGSTRVCGS 173

Query Match	67.5%	Score 2852;	DB 21;	Length 758;
Best Local Similarity	67.5%;	Pred. No. 1.2e-260;		
Matches 530;	Conservative 94;	Mismatches 133;	Indels 28;	Gaps 9

[illegible][illegible]

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:02:11 ; Search time 28 seconds
(without alignments)

1161.337 Million cell updates/sec

Title: US-09-675-208-5

Perfect score: 4225

Sequence: 1 MAPSFKNKGNNVVSFDGLND.....QVWPIDSSSGISVIEYLF 784

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	115.5	2.7	902	YC47_SCHPO	O14053 schizosacch
2	107	2.5	1174	CIEB_BACTA	O03745 bacillus th
3	106	2.5	1723	AIM1_HUMAN	O944k1 homo sapien
4	105	2.5	730	GLGB_HAEIN	P45177 haemophilus
5	102.5	2.4	564	HEMA_IARUD	P19700 influenza a
6	101.5	2.4	478	IRBK_SALCH	O00330 salmonella
7	101	2.4	514	VG64_HSV1	O00156 ictaluriid h
8	100.5	2.4	712	IRAL_HUMAN	P51617 homo sapien
9	100	2.4	1550	GLTB_SYNY3	P55037 synochocyst
10	99	2.3	770	GLGB_SYNY3	P52981 synochocyst
11	99	2.3	1137	CYGL_CAEEL	O09435 caenorhabd
12	98	2.3	306	YCBJ_BACSU	P42242 bacillus su
13	97.5	2.3	1095	PIPA_DROME	P13217 drosophila
14	97	2.3	520	HMC5_CRIGR	P13704 cricetus
15	97	2.3	520	HMC5_RAT	P17425 rattus norv
16	97	2.3	859	ENV_ETAV2	P22428 equine infe
17	97	2.3	1002	TAGA_VIBCH	P24019 vibrio chol
18	97	2.3	3061	POLG_PVYHU	O02963 p genome po
19	96.5	2.3	444	YIHL_YEAST	P40512 saccharomyc
20	96.5	2.3	631	DMH1_PARDE	P12293 saccharomyc
21	96	2.3	460	ALN_YEAST	P33293 paracoccus
22	96	2.3	520	HMC5_HUMAN	O01581 homo sapien
23	95.5	2.3	564	HEMA_IAGRE	P16698 influenza a
24	95.5	2.3	346	RUVB_MCMV	O9a398 maize chlor
25	95	2.2	705	CATE_RHIME	O9a576 rhizodium m
26	95	2.2	791	LON_CAMJE	O63300 campylobact
27	94.5	2.2	733	AGAI_PEDPE	P43467 pedicoccus
28	94.5	2.2	971	NAC1_RAT	O01728 rattus norv
29	94.5	2.2	1072	CARB_THETN	O8dbk0 thermoeaer
30	94	2.2	521	IMAL_XENLA	P52170 xenopus lae
31	94	2.2	563	CHIA_SERMA	P07254 serratalia ma
32	94	2.2	843	PULA_THEMA	O33840 thermotoga

34	93	2.2	695	1	EEG1_SYNY3	P28371 synochocyst
35	93	2.2	859	1	ENV_ETAV1	P22427 equine infe
36	93	2.2	1047	1	ANPB_BOVIN	P46197 bos taurus
37	93	2.2	3988	1	POLG_BVDVN	P19711 bovine vira
38	92.5	2.2	400	1	EPFU_HERNU	P42477 herpesosiph
39	92.5	2.2	564	1	HEMA_IASE2	P19701 influenza a
40	92.5	2.2	646	1	DHSA_CAEEL	O09508 caenorhabd
41	92.5	2.2	1356	1	HER1_PODAN	O00808 podospora a
42	92.5	2.2	1431	1	DAPK_HUMAN	P53355 homo sapien
43	92	2.2	859	1	ENV_ETAVC	P32541 equine infe
44	92	2.2	966	1	M172_HUMAN	O14596 homo sapien
45	92	2.2	1065	1	CARB_CLOAB	O97f13 clostridium

ALIGNMENTS

```

RESULT 1
YC47_SCHPO          STANDARD:      PRT:      902 AA.
ID                  YC47_SCHPO
AC                  O14053:
DT                  15-JUL-1998 (Rel. 36, Last Created)
DT                  15-JUL-1998 (Rel. 36, Last sequence update)
DT                  15-JUN-2002 (Rel. 41, Last annotation update)
DE                  Hypothetical WD-repeat protein Cl672.07 in chromosome III.
GN                  SPC1672.07.
OS                  Schizosaccharomyces pombe (Fission yeast).
OC                  Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC                  Schizosaccharomycetales; Schizosaccharomycetaceae;
OC                  Schizosaccharomycetes.
OX                  NCBI_TaxID=4896;
RN                  [1]
RP                  SEQUENCE FROM N.A.
RC                  STRAIN=972;
RX                  MEDLINE=21848401; PubMed=11859360;
RA                  Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA                  Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA                  Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA                  Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA                  Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson C.,
RA                  Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA                  James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA                  Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA                  Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,
RA                  Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA                  Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA                  Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA                  Woodward J., Volkert G., Aert R., Robben J., Grynopre B.,
RA                  Welljens I., Vanhulst E., Rieger M., Schaefer M., Mueller-Auer S.,
RA                  Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA                  Borzym K., Langer I., Beck A., Leirich E., Reinhardt R., Pohl T.M.,
RA                  Eger P., Zimmermann W., Wedler H., Mambuti R., Purnelle B.,
RA                  Goffeau A., Cadieu E., Dreano S., Gloux S., Leleure V., Moutier S.,
RA                  Gallbert F., Aves S.J., Xiang Z., Hunt G., Moore K., Hurst S.M.,
RA                  Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,
RA                  Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA                  Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA                  Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA                  Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RT                  "The genome sequence of Schizosaccharomyces pombe."
RL                  Nature 415:871-880(2002).
-i- SIMILARITY: CONTAINS 11 WD REPEATS (TRP-ASP DOMAINS).
- This SWISS-PROT entry is copyright. It is produced through a collaboration
- between the Swiss Institute of Bioinformatics and the EMBL outstation
- the European Bioinformatics Institute. There are no restrictions on its
- use by non-profit institutions as long as its content is in no way
- modified and this statement is not removed. Usage by and for commercial
- entities requires a license agreement (See http://www.isb-sib.ch/announce/
- or send an email to license@sib-sib.ch)
EMBL: AL031324; CNA20445.1; -
InterPro: IPR001680; WD40.

```



```

Db 650 FCLDEKRELSKVAKKRLSDERNLLQDPNFRGINRQDPRGRSGSTDTIIGCGDDVFKEN 709
Qy 642 PISIEGV--KTFALYLYQAKLLSK-----PSODLDIALDPF--EPELIT 683
Db 710 YVTLCTEPECEPTLYIXQ--KIDESKLVNRYQLRGYIEDSODLEIYIRINAKHEIYN 767
Qy 684 VSPVTKLIQTSLHPAIGLVNMLNTSGAIOGVYDDDL-----SVEIG 727
Db 768 VGTGSLWPLSYE-SPIGSGEPNR--CVPHEMNPDLDSCSDRGEKCAHSHHSFLDID 824
Qy 728 VNGGEMRVFASKRPACRIDEDVNG 753
Db 825 V-GCTDLQ-----EDLG 835

RESULT 3
ATM1_HUMAN STANDARD: PRT: 1723 AA.
AC 09YAK1; O00296;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Absent in melanoma 1 protein.
GN ATM1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97250519; PubMed=9096375;
RA Ray M.E., Mistow G., Su Y.A., Meltzer P.S., Trent J.M.;
RT "ATM1, a novel non-lens member of the betagamma-crystallin
RT superfamily, is associated with the control of tumorigenicity in human
RT malignant melanoma."
RT Proc. Natl. Acad. Sci. U.S.A. 94:3229-3234(1997).
CC -1- FUNCTION: MAY FUNCTION AS SUPPRESSOR OF MALIGNANT MELANOMA. IT MAY
CC EXERT ITS EFFECTS THROUGH INTERACTIONS WITH THE CYTOSKELETON.
CC -1- SIMILARITY: BELONGS TO THE BETA/GAMMA-CRYSTALLIN FAMILY. CONTAINS
CC 6 TANDEM REPEATS OF A BETA/GAMMA-TYPE CRYSTALLIN DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 RICHIN B-TYPE LECTIN DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
DR EMBL: U83116; AB53792.1; -
DR EMBL: U83115; AB53791.1; -
DR HSP: P02526; IGAM.
DR Genew: HGNC:356; ATM1.
DR MIM: 601797; -
DR InterPro: IPR001064; Crystallin.
DR InterPro: IPR000772; Richn_B_lectin.
DR Pfam: PF00030; crystal1; 6.
DR Pfam: PF00652; Richn_B_lectin; 3.
DR PRINTS: PRO1367; BGCRCRYSTALLIN.
DR SMART: SM00458; RICHIN; 1.
DR SMART: SM00247; XTALBP; 6.
DR PROSITE: PS00325; CRYSTALLIN_BETAGAMMA; 7.
DR PROSITE: PS50231; RICHIN_B-LECTIN; 1.
DR Repeat: Lectin.
FT DOMAIN 1022 1119 CRYSTALLIN BETA/GAMMA-LIKE 1.
FT DOMAIN 1123 1206 CRYSTALLIN BETA/GAMMA-LIKE 2.
FT DOMAIN 1207 1318 CRYSTALLIN BETA/GAMMA-LIKE 3.
FT DOMAIN 1319 1404 CRYSTALLIN BETA/GAMMA-LIKE 4.
FT DOMAIN 1405 1487 CRYSTALLIN BETA/GAMMA-LIKE 5.
FT DOMAIN 1502 1584 CRYSTALLIN BETA/GAMMA-LIKE 6.
FT DOMAIN 1586 1719 RICHIN B-TYPE LECTIN.

```

```

SQ SEQUENCE 1723 AA; 188646 MW; 7E50F681A627FB09 CRC64;
Query Match 2.5%; Score 106; DB 1; Length 1723;
Best Local Similarity 17.8%; Pred. No. 5.3;
Matches 156; Conservative 127; Mismatches 301; Indels 292; Gaps 41;
Qy 13 VSPDGLNDSSPEFALDGSDFITVNGHSFLSDVPENIVASPSPTSIDKSPVSGCFVGF-D 71
Db 996 ISFGSGS-LSDMITLRGS-----YQNKLNPRGKVIYISEPVSKEICIEVFSD 1042
Qy 72 ASEPDSSRHVVSIGKLDKIDREMSIFREKVMWTHWVGRNGDLE-----SEFOIYLEKS 125
Db 1043 IODCSSWSLSPIVILKVGWCIWEQPNFECHSIPLIEGELELSGLMGIEDILERHEBA 1102
Qy 126 DSGRPVFLPLIVEGFRFRTSIQPGDDPFVDCVVESSSKVVDASFRSMYLHAAGDP--- 182
Db 1103 ESDK-----PVIIGSIRIVODYRVSHDLFTEPBGILGILSSYF-----DTEEM 1147
Qy 183 --FALVKEAMKIVRTHLGFRLIEKTPPGI-----VDKFGWCWDAFYLVNRP-- 229
Db 1148 QGFGVWQKTCG-MKVHWGTLWILYEEPGFQGVPILEPGEPDLSFDWTFAAYIGSMRPLK 1206
Qy 230 -----OGVIEGVHVLVDGCGPGLVLIDGW-----OSIGHSDPIRK 267
Db 1207 MGRKVEFPPTDKVYVYERKPFEGKC---VELETGMCSEFVMEGGETEATGDDHLPFS 1262
Qy 268 EGMNQIVAGEQMPCKRLKQENYKFRDYVNPKATGRAGOKMKAFIDLKKEKTEVHV 327
Db 1263 VGSMAKVLKGIWVA-----YEKP-----GFTGHQYLEE--GEYRQ--- 1295
Qy 328 YVWHALGCGWGLRPQGLPPEARVIOPLVSPGLQWTMEDLAVDKIVLKVGLVPEKAE 387
Db 1296 --WKA-----WGVTNGELQSL-----RPLG-----DFSNAHIMVSEKNFGSGSS 1335
Qy 388 EMEGIEHAHLEKVGIDGVKIDVILHLEMLCEDY-----GGRVLDKAVYKAKTKSINKH 441
Db 1336 IYVLGIIVANLKETGY-GVYQTSINVLGVAVAYENPDFGEQYIIDKGFYTFEDWGKN 1394
Qy 442 FPGNGVIAEMENCDNMFLETGTAISLGRGDDFWCIPDPSGDPNGRTWLOGCHVHNCANS 501
Db 1395 YK-----ISSVQPI-----CLDSFTGPPRR-----NOIHLESEP 1423
Qy 502 LMMGNFIHPDMWFOSTHCA--AFHAASRAISGGIYVSDS-----VGKHNFDLL 550
Db 1424 QVQGHG-----QSFEETTSQIDDSFTKSCRYSGGSWVYVDGENTFGNGLVEEGHYPL 1478
Qy 551 KKLVLDPDGSILRSEYALPTRDCLFEDPL-----HNGETMLKIWNLNKFT 595
Db 1479 SAMGCPPGATFKS---LRFIDVEFSEPTIILFEREDFKKIELNAETV---NLRSL- 1529
Qy 596 GVIAGFNCG-----GGGW-----GRETRRNCQFSQ 620
Db 1530 ----GFNTQIRSVQVIGIWTYVEGSGYRGROFLSPAEVPMWYEESSGROGSLRPFQ 1585
Qy 621 YSKRVTSKTNPKDIEWHSSENPISIEGVKTF-----ALYLYQAKLLSKPSQ 668
Db 1586 --KRIYFRLNKATGLFMTNG-NLEDKLKLRIQWEDVGADQDIWITYQ--EGCICKRIAE 1641
Qy 669 DLDIALDPEPELITVSPVTKLIQTSLHPAIGLVNMLNTSGAIOGVYDDDLSSVEIGV 728
Db 1642 DCLLT-----IVSGLTSGSKLGLALDQNASQFWSL 1673
Qy 729 KGGEMRVFASKRP-ACRIDEDVGFKKYDQDMV 763
Db 1674 KSDG--RIYSKLPNLVLDIKG---GTQYDQNHIL 1704

RESULT 4
GGB_HAEIN
ID GGB_HAEIN STANDARD: PRT: 730 AA.
AC P45177;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (Glycogen branching
DE enzyme).
GN GlgB OR H11357.
OS Haemophilus influenzae.
OC Bacteria: Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
NCBI_Taxid=727;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / K20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierulff A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shrivley R., Liu L.-I., Glodok A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uitterlind T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geohegan N.S.M.,
RA Genlin C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RT Science 269:496-512(1995).
CC -!- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
CC glycogen.
CC -!- PATHWAY: glycogen biosynthesis; third step.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U32815; AAC23004.1; -
DR TIGR: H11357; -
DR InterPro: IPR000461; Alpha-amylase.
DR InterPro: IPR004193; Isoamylase_N.
DR Pfam: PF00128; alpha-amylase_1.
DR Pfam: PF02922; Isoamylase_N.1.
KW Glycogen biosynthesis; transferase; glycosyltransferase;
KW Complete proteome.
FT ACT_SITE 405 BY SIMILARITY.
FT ACT_SITE 458 BY SIMILARITY.
FT ACT_SITE 526 BY SIMILARITY.
FT ACT_SITE 526 BY SIMILARITY.
SQ SEQUENCE 730 AA; 83820 MW; 5B957531F53769A CRC64;
Query Match 2.5%; Score 105; DB 1; Length 730;
Best Local Similarity 17.5%; Pred. No. 1.8;
Matches 117; Conservative 75; Mismatches 203; Indels 272; Gaps 29;

```

```

DB 324 -----
DB 349 EARYIOPVLSPIQMTMEDLAVDKIVLHKVGLVPEPAEEMEGELAH ----- 346
DB 347 -----
DB 347 -----GL-----VAFDGALY-----EHEDPRRGYHODMWTLYVNGRNE 381
DB 397 -----LEKVIDGKIDVITHL-----EMLCEDYGGRYDL-AKAYKA 433
DB 382 VKNPFLSNALYWERFEVDGIRDAVASMTYROYSRABEGWIRNOJGRNLEAIEFLKH 441
DB 434 MTSINKHFKNGVYASMEHCNDPMLCTAISLGRVDDFWCTDBSGDPNGTFWLOGCH 493
DB 442 TNWKI--HSMAGASIAESTSFAGV-THPSRNGGLGFNF-----KWNMG-- 484
DB 494 MIVCANDSLWNGNFIHPDWMFQSTHRCAFHAASRAISGSPITYVSDVKHNFLLKL 553
DB 485 -----MMNDTL-----AYMKLDPIYROYHNKMKTFGMYOY 515
DB 554 VLPDGSILRSEYVALPTROCLFEDPLHN-----GETMLKIVNLNKFVIGAF--- 601
DB 516 -----SENFVPLS-----HDEVHGKYSILGKMPGDIWQKFAIRAYGYFMWGPCK 563
DB 602 -----NCQ-----GGGWCRTTR-----NOCESQYSKRVTSKT 629
DB 564 KLFMGNEFAQGRWVVEESLDWFLDENIGGMHKGVLKLVLDNIYOKNRPFLDKN 623
DB 630 NPKDIEW 636
DB 624 SPEGFDW 630
RESULT 5
HEMA_IARUD
ID HEMA_IARUD STANDARD: PRT: 564 AA.
AC P19700;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemagglutinin precursor [Contains: Hemagglutinin HAI chain;
DE Hemagglutinin HAZ chain].
GN HA
OS Influenza A virus (strain A/Ruddy Turnstone/New Jersey/47/85).
OC Viruses: ssRNA negative-strand viruses: Orthomyxoviridae;
OC Influenza A viruses; Influenzavirus A.
OX NCBI_Taxid=11458;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89204912; PubMed=2705304;
RA Donis R.O., Bean W.J., Kawabata Y., Webster R.G.;
RA "Distinct lineages of Influenza virus H4 hemagglutinin genes in
RT different regions of the world.";
RT Virology 169:408-417(1989).
CC -!- FUNCTION: HEMAGGLUTININ IS RESPONSIBLE FOR ATTACHING THE VIRUS TO
CC CELL RECEPTORS AND FOR INITIATING INFECTION.
CC -!- SUBUNIT: HOMOTRIMER. EACH OF THE MONOMERS IS FORMED BY TWO CHAINS
CC (HA1 AND HA2) LINKED BY A DISULFIDE BOND.
CC -!- SIMILARITY: BELONGS TO THE INFLUENZA HEMAGGLUTININ FAMILY.
DR PIR: G34214; HMIY7.
DR HSSP: P03437; 2VIU.
DR InterPro: IPR001364; Hemagglutn.
DR Pfam: PF00509; Hemagglutinin; 1.
DR PRINTS: PR00329; HEMAGGLUTN12.
DR ProDom: PD000225; Hemagglutn; 1.
KW Envelope protein; Hemagglutinin; glycoprotein; signal.
FT SIGNAL 1
FT CHAIN 17
FT CHAIN 17 342
FT CHAIN 344 564
FT CARBOHYD 18 18 HEMAGGLUTININ HAI CHAIN.
FT CARBOHYD 34 34 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 178 178 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 310 310 N-LINKED (GLCNAc. . .) (POTENTIAL).
FT CARBOHYD 497 497 N-LINKED (GLCNAc. . .) (POTENTIAL).

```

SQ SEQUENCE 564 AA: 63271 MW: D4010C2548AE3EFC CRC64;
 Query Match 2.4%; Score 102.5; DB 1; Length 564;
 Best Local Similarity 18.6%; Pred. No. 2;
 Matches 110; Conservative 61; Mismatches 186; Indels 233; Gaps 26;

```

OY 7 NGSNVVSEFDGLND-----MSSPFAIDGSDFTVNGHSLSDVPENIVASPSPTSIDKS 60
DB 80 NGAISPGCDHNLNGAMBVDIERPTAVD-----TCYFVDPD-----YQSIRSI 123
OY 61 PVSVCCEFGDASEPDSRRHVSIGK---LKDIRFMSIFEFKXWVTHHWGR----- 108
DB 124 LANNKRF-EFIAEEFQWMTVKONGKSGACKRANVDFER-RLNWLTKSGNAYPLONLTK 181
OY 109 -NGGDL-----ESETQIVLEKSDGR-----PYVFLPIVEGP-- 141
DB 182 VANGDYARLYTNGVHHPTSDTQTNLYKNNPGRVYVSTQTSQSVYVPIGSRPLVNGGSG 241
OY 142 ---FRTSIOPGD-----DDFVDCVESGSS 163
DB 242 RISFYWTIVEPDLLVFNTIGNLIAPRGHYKINSOKKSTILMTAVPIGSCVSKCHTNRGS 301
OY 164 KYVDASFNSMLYLHAGDDPFALVKEAMKIVRHLGLTFLLEKTPPGIYDKRGCTWDAF 223
DB 302 IYTRPFQNIISISIGDCRYVKOSLKLAT---GMRIPIPKATRGLE----- 346
OY 224 YLTVHPQVIEGVRLVLDGCPFGVLIDDGQSIQHDSDPTKEGMNQTVAGEQMPRL 283
DB 347 -----GAIAG-----FIENGNGL----- 360
OY 284 LKFOENYKRDYVNPRAKTPRAGQKMAFIDELKGEFTVEHYVYHVALCGYWGGLRQ 343
DB 361 --IDGWYGR-HQNEGTCTADLKTQAIDQINGKLNRL----- 398
OY 344 VGLPRLARIQVLSGLQMTMEDLAVDKIVLHKVGLVPEREAEEMEGHAELEKVGID 403
DB 399 -----IKETNEK-----YHQI-----EKEFEQVEIRIDLEKY-VE 428
OY 404 GVKIDV-----HLEMLCEDYGRVDLAKAYKAMTKSI-----NKHFKNGVISMEEH 453
DB 429 DTKIDLMSTNAELVALLEMH--TIDVTSEMKKLEFERRRLQRENAEDKGNCFEIFRD 486
OY 454 CNDFMFLTEAISLGRVGDFFMCTDPSGDPNGTFMLQGGHMYHACND-SLWM 504
DB 487 CONNCT--ESTRNGTYDHDYRDEAI---NNRFOIQGVKLTGQYKDIILMI 532
  
```

RESULT 6
 REBK_SALCH STANDARD; PRT; 478 AA.
 AC 000330;
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 01-OCT-1994 (Rel. 30, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphomannomutase (EC 5.4.2.8) (PMM).
 GN MANB OR REBK.
 OS Salmonella cholerae-suis (Salmonella enterica).
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Salmonella.
 OX NCBI_TaxId=591;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=M67;
 RX MEDLINE=92349966; PubMed=1379320;
 RA Brown P.K., Romana L.K., Reeves P.R.:
 RT "Molecular analysis of the rfb gene cluster of Salmonella serovar
 RT muenchen (strain M67): the genetic basis of the polymorphism between
 RT groups C2 and B.";
 RL MOL. MICROBIOL. 6:1385-1394(1992).
 CC -I- FUNCTION: INVOLVED IN GDP-MANNOSE BIOSYNTHESIS WHICH SERVES AS
 CC THE ACTIVATED SUGAR NUCLEOTIDE PRECURSOR FOR MANNOSE RESIDUES
 CC IN CELL SURFACE POLYSACCHARIDES. THIS ENZYME PARTICIPATES IN
 CC SYNTHESIS OF THE LPS GROUP C2 O ANTIGEN.

CC -I- CATALYTIC ACTIVITY: D-mannose 1-phosphate -> D-mannose 6-phosphate.
 CC -I- PATHWAY: GDP-MANNOSE BIOSYNTHESIS WITHIN THE O ANTIGEN
 CC BIOSYNTHESIS PATHWAY OF LIPOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -I- SIMILARITY: BELONGS TO THE PHOSPHOHEXOSE MUTASES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```

DR EMBL, X61917; CAA43916.1; .
DR PIR, S22622; S22622.
DR InterPro: IPR001485; PG/PMM_mutase.
DR Pfam: PF00408; PGM_PMM; 1.
DR Pfam: PF02878; PGM_PMM_I; 1.
DR Pfam: PF02879; PGM_PMM_II; 1.
DR Pfam: PF02880; PGM_PMM_III; 1.
DR PROSITE: PS00710; PGM_PMM; 1.
KW Lipopolysaccharide biosynthesis; Isomerase; Transmembrane;
KW Phosphorylation.
FT TRANSMEM 30 46
FT TRANSMEH 265 284
FT ACT_SITE 111 111
FT FT
SQ SEQUENCE 478 AA: 52813 MW: 9825B885083893 CRC64;
  
```

Query Match 2.4%; Score 101.5; DB 1; Length 478;
 Best Local Similarity 21.8%; Pred. No. 1.9;
 Matches 59; Conservative 30; Mismatches 79; Indels 103; Gaps 14;

```

OY 349 EARVIOPVLSPLQMTMEDLAVDKIVLHKVGLVPEREAEEMEGHAELEKVGID 408
DB 141 DASFMQPKLE---QLTISTILANRYLTSLTFMPLKRRKRGITDH----- 185
OY 409 VYHLEMLCEDYGRVDLAKAYKAMTKSINKHFGKNGVYASNEHCNDFMFLTEAISLG 468
DB 186 -----SSAGR-DLVKTLFLKML-----GATVYSLARSDEFVPIDEAVS-- 222
OY 469 RVGDD-----FMCIDPSG-----DRNGTFWLOG-----GIM----- 494
DB 223 --EDDRNKAITWAKKYQLDALIFSTDGDRPLADEYGN-WLFGDILGLCSLELADAV 279
OY 495 ---VHCANDSLWKGNTL-HPDMDMFQSTHPCAFHAASTRASISGPIYSDSVGKHNPDLL 550
DB 280 AIPVSC-NSTISSGNFFKHYVERTKISPIYIAFAKLSA-----NYNCT 322
  
```

RESULT 7
 VG64_HSV11 STANDARD; PRT; 514 AA.
 AC 000156;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 01-DEC-1992 (Rel. 24, Last annotation update)
 DE Hypothetical gene 64 protein.
 GN 64.
 OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Ictalurid herpes-like viruses.
 OX NCBI_TaxId=10401;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=Auburn 1;
 RX MEDLINE=92087490; PubMed=1727613;
 RA Davison A.O.:
 RT "Channel catfish virus: a new type of herpesvirus."

RL Virology 186:9-14(1992).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: M75136; AAA8168.1; -
 DR PIR: A36793; A36793.
 KW Hypothetical protein.
 SQ SEQUENCE 514 AA; 58198 MW; 88E7302A142AD29C CRC64;
 Query Match 2.4%; Score 101; DB 1; Length 514;
 Best Local Similarity 20.3%; Pred. No. 2.3; Mismatches 177; Indels 172; Gaps 28;
 Matches 108; Conservative 75; Mismatches 177; Indels 172; Gaps 28;
 QY 33 TVNGHSLSDVPENIVASPSPTSIDSQSPVSGFVG-----FDASEPDSRHVST 83
 DB 44 TLEGEDELEGIPQELINLVLYEKMKSTKSELGFTVYALMTLGIFEDR-----SL 96
 QY 84 GLIKDIREHSIRPEKVMTHWYGRNGDLESTQVILEKSDSGRPYFL--PIVeg 140
 DB 97 KRLVDCVGMERENMLTRVFFGYATGD---DALIGLDTWPAFVLLARFVIPA 151
 QY 141 PERTSPDDEFDVDCV-ESGSS-KVYDASFRSMLYH-----AGDDP-----FAL 185
 DB 152 -----QLDSNFCVLCMESDHSLATADTRFTLICDHTEASVYAGDPCSAEYTA 205
 QY 186 VDEAKIVTHLGTFRLEKTPPGIVDKFGWCTWDAEYLVHPQGVIGVRHLVDGCG 245
 DB 206 VKKTQI-DLSLDDF--HEVARG-----DNVREL-----232
 QY 246 PGVYLIDGMOISGHSDPITKEG-----MNQTVAGEOMPCLRKQENYKFR 293
 DB 233 -----AKKRDITLTTRKNGSINNFTYALNPARRETCADRNESSTVIFOKLREV 281
 QY 294 DYVNPATGPRAQOKMAFIDELKEFTEVHVYHVALCGYGGRLQVPGPLPARYV 353
 DB 282 NMNINLTNP---KVLNFEALFNVFVAHH-----RRKTI 315
 QY 354 QVYLSGL-----QMTMEDLVADKIVLKV-----GLVPEKAEEMYGCLA 395
 DB 316 RERLSDGLDMKGRRAHMLFTKRLTAADKMLTITIEFYAAAPVRYRNHKNYMA 375
 QY 396 -----HLEKVGI--DGV-----KIDVHLEMLCEDYGRVLDKAYYKATKTSINK 441
 DB 376 HAVIFPKHESIGFTMEGITAFOYQLNQVDLREVQDI--RSDIIPITTN--ENFNRL 431
 QY 442 FKNGVYIASMEHCNDEMLGTETASLGRVDDFWCTDPSGD--PNCFTWLOG 491
 DB 432 TELEVIANI--CT-FIFHNKTIKLH-GD--APDLNIDELPDGLYLFNG 476
 RESULT 8
 ITAL_HUMAN STANDARD; PRT; 712 AA.
 ID ITAL_HUMAN P51617;
 AC P51617;
 DT 01-OCT-1996 (rel. 34, Last Created)
 DT 01-OCT-1996 (rel. 34, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Interleukin-1 receptor-associated kinase 1 (EC 2.7.1.-) (IRAK-1).
 GN IRAK1 OR IRAK.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE=96180673; PubMed=8599092;
 RA Cao Z., Henzel W.J., Gao X.;

RT "IRAK: a kinase associated with the interleukin-1 receptor."
 RL Science 271:1128-1131(1996).
 RN [2]
 RA SEQUENCE FROM N.A.
 RA Platzer M., Bauer D., Drescher B.;
 RN Submitted (MAR-1995) to the EMBL/GenBank/DBJ databases.
 RP [3]
 RP SEQUENCE FROM N.A.
 RA Reichwald K., Kioschis P., Rosenthal A., Platzer M.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 CC -FUNCTION: INVOLVED IN IL-1 PATHWAY. THIS KINASE ASSOCIATES WITH
 CC THE IL-1 RECEPTOR IL-1R-1. THIS ASSOCIATION IS RAPID AND IL-1
 CC DEPENDENT.
 CC -TISSUE SPECIFICITY: SEEMS TO BE UBQUITOUS, ALTHOUGH PRESENT IN
 CC SMALL AMOUNTS.
 CC -PTM: AN EXTENSIVE PHOSPHORYLATION OF IRAK OCCURS AFTER ITS
 CC ASSOCIATION WITH IL-1R-1. THIS STEP COULD BE LINKED TO THE
 CC ACTIVATION OF THE KINASE.
 CC -SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
 CC PELLE SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: U76191; AAC1949.1; -
 DR EMBL: U52112; AAC1752.1; -
 DR EMBL: AF030876; AAC08756.1; -
 DR EMBL: AF031075; AAF21636.1; -
 DR Genew; HGNC:6112; IRAK1.
 DR MIM: 300283; -
 DR Interpro: IPR000488; Death.
 DR Interpro: IPR000719; Euk_pkinase.
 DR Interpro: IPR002290; Ser_thr_pkinase.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF00531; death; 1.
 DR Prodom: PD000001; Euk_pkinase; 1.
 DR Prosite: PS00107; PROTEIN_KINASE_ATP; 1.
 DR Prosite: PS00108; PROTEIN_KINASE_DOM; 1.
 DR Prosite: PS50011; PROTEIN_KINASE_ST; 1.
 DR Transferase: Serine/threonine-protein kinase; ATP-binding.
 KW DOMAIN 212 521
 FT NP_BIND 218 226 ATP (BY SIMILARITY).
 FT BINDING 239 239 ATP (BY SIMILARITY).
 FT ACT_SITE 338 338 BY SIMILARITY.
 FT CONFLICT 196 196 F -> S (IN REF. 1).
 FT CONFLICT 332 332 S -> L (IN REF. 1).
 SQ SEQUENCE 712 AA; 76536 MW; A7ADEB75D3A3981D CRC64;
 Query Match 2.4%; Score 100.5; DB 1; Length 712;
 Best Local Similarity 27.0%; Pred. No. 4.1; Mismatches 61; Indels 25; Gaps 6;
 Matches 38; Conservative 17; Mismatches 61; Indels 25; Gaps 6;
 QY 116 ETQIVILEKSDSGRPYFLPIVIEGPFRTSIQDDEFDVDCV--ESGSSKYVDASFRM 173
 DB 184 ESSVSILQ---GARPPPCPLCE-----ISRGTHNFSBELKIGCGFGCVYRAVRMT 234
 QY 174 LYL-----HAGDDPFLVDEAKIVTHLCTFRLEKTPPGIVDKFGWCTWDAEYLVH 228
 DB 235 VYAVKRLKENADEWPAVQSPFLTEVEQDSRFR-----HPNIVDAGCAONGFYCLVY 288
 QY 229 ---PGVIGVRLVVDGCGCP 246
 DB 289 GFLPNSLEDRLHCTQACPP 309
 RESULT 9
 GLTB_SYNY3 STANDARD; PRT; 1550 AA.
 ID GLTB_SYNY3

```

AC P55037.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ferredoxin-dependent glutamate synthase 1 (EC 1.4.7.1) (Fd-GOGAT).
GN GLTB OR SLI1502.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95244836; PubMed=7727752;
RA Navarro F., Chavez S., Candau P., Florencio F.J.;
RT "Existence of two ferredoxin-glutamate synthases in the
RT cyanobacterium Synchocystis sp. PCC 6803. Isolation and insertional
RT inactivation of gltb and glts genes.";
RL Plant Mol. Biol. 27:753-767(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hoshino T., Matsuno A., Muraki A., Nakazaki N., Nario K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-116(1996).
CC -1- CATALYTIC ACTIVITY: 2 L-glutamate + 2 oxidized ferredoxin = L-
CC glutamine + 2-oxoglutarate + 2 reduced ferredoxin.
CC -1- COFACTOR: BINDS A 3FE-4S CLUSTER; FAD AND FMN.
CC -1- PATHWAY: GLUTAMINE SYNTHETASE/GOGAT PATHWAY WHICH IS INVOLVED
CC IN THE ASSIMILATION OF AMMONIA.
CC -1- SIMILARITY: TO OTHER GLUTAMATE SYNTHASES.
CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-21 IS THE INITIATOR.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC DR EMBL: X80485; CAA56552.1; -
CC DR EMBL: D90902; BAA17018.1; -
CC DR InterPro: IPR002489; DUF14.
CC DR InterPro: IPR002932; Glu_synthase.
CC DR Pfam: PF01493; DUF14; 1.
CC DR Pfam: PF01645; Glu_synthase; 1.
CC DR Oxidoreductase: Iron-sulfur; 3Fe-4S; Flavoprotein; FAD; FMN;
CC KW Glutamate biosynthesis; Complete proteome.
CC FT DOMAIN 43 393
CC FT NP_BIND 1097 1150
CC FT METAL 1150 1150 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
CC FT METAL 1156 1156 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
CC FT METAL 1161 1161 IRON-SULFUR (3FE-4S) (BY SIMILARITY).
CC SO SEQUENCE 1550 AA; 169071 MW; 07AA64DB32296943 CRC64;
Query Match 2.4%; Score 100; DB 1; Length 1550;
Best Local Similarity 19.9%; Pred. No. 14;
Matches 136; Conservative 89; Mismatches 224; Indels 234; Gaps 38;
QY 120 VLEKSDSRPVFLPIYEGPFRS-----IQGDDDFDVCEVSSSKVAVDSFMSLY 175
DB 406 LVIMASEAG---VLP---EPERAAKGRLOPGRMFLVD--MEQGR-ITAEDEIK---- 451
QY 176 LHAQDDPFLVAKMEKIVRTH-----LGTFRLEKTPGIVD-----KFGM 217
DB 452 -----GEIVSQHYGEWLANLAKSLDGLSPGANVPGTDAISLGRQNAFXY 497
QY 218 CTWDAFYLTVPQGVIEGVRLHVDGCGPGGLVLIDG-----WQSIGHSDPI 265

```

```

DB 498 -FEEELRILLDMG-RDGEAIGSMGADTPLAVLSDBKRLLYNYEQLEAQTNPDISI 555
QY 266 TRE---GMNQVAGE-----QMPCTL-----LKEQENKRDVYNPKA 300
DB 556 REEITTSATTGGGCGNLDPRESCRLIELKTPILTNEDLAKLALDDDEKSTVLDL 615
QY 301 TGPRAQGMKAFIDELGGEFK-----YVEHYVMHALCGYWG----- 338
DB 616 FDPNGEGAGLKTALDNLFTADQALISOGANLITLSDQVSAEKAL-PAALLAVSGLHHL 674
QY 339 ---GLRPVPGIPPEARVIOPVLSPLQMTMEDLAVDKIYLVKPPKAEEMEGHLA 395
DB 675 IANGSRKV-GL-----VLESGEPREVNHFV--LLGYCGAINPYLAETLDGMIA 723
QY 396 HLEKVGIDVKIDVTHLEMLCEDYGGRVDLKAVYKAKTSINHFNGNGVIASMECN 455
DB 724 -----EGLLVNVH--KTAC-----KNIKATKGVITVASKIGISTQSTRG 764
QY 456 DEMPLGTEAISLGR-VGDDFWCTDPSGDPNGTFLWQGHMVHCANDSLMGNFIHPDWM 514
DB 765 AQIF--EAVGLNQGVIVDEIFCRTSSR-----IQGSDLVIAQEAIR----- 804
QY 515 FOSTHPCAAFHAAISGSPFIYVSDSVKANF-----DLKRLVLPDGTI 560
DB 805 -----HGHAFAPRPGDLHTLDVGEYQWRKDGEHLFPOTIHLQRAVEGNYE 854
QY 561 LRSEYVAL-----PTROCL-FED-----PLHNGETMLKIMNLNK-F-TGVICAFNCG 605
DB 855 LKQYALVNEONOKFFTLRGLDPQDRESIPLEVEPIEAL--MKRFETGAMST----- 907
QY 606 GWCMEETRRNOCFSOYSKRVTSKTN-----PKDIEHSGENPISIEGVKTF----- 952
DB 908 GSISNEAHESLIAM--NNGKSKMTGEGCEDEPERFTWINDGDSKNSAIKQVASGREYV 965
QY 653 --LYIQAKKIL-----SKPSQ 668
DB 966 TSLVLSQAKEIOIKMAQKAPGE 988
RESULT 10
GLGB_SYNY3
ID GLGB_SYNY3 STANDARD: PRT: 770 AA.
AC P52981;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE 1,4-alpha-glucan branching enzyme (EC 2.4.1.18) (glycogen branching
DE enzyme).
GN GLGB OR SLI0158.
OS Synchocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synchocystis.
OX NCBI_Taxid=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugiyama M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synchocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 64% to 92% of the genome.";
RL DNA Res. 2:153-166(1995).
CC -1- CATALYTIC ACTIVITY: Formation of 1,6-glucosidic linkages of
CC glycogen.
CC -1- PATHWAY: Glycogen biosynthesis; third step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO FAMILY 13 OF GLYCOSYL HYDROLASES, ALSO
CC KNOWN AS THE ALPHA-AMYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```


OY 298 PRATGRAGQKMKAKIDELKGEFKEVHYVWALCGYWGRLPOVPLPEARVQPV 357
 DB 814 PRVENICSOQMKGLVS-----KOKTNLMHVF--NMDEXSTLSEET-----EERT----- 857
 OY 358 SEPOLMTMEDLAVDKVLKVLGVPEPEKAEEMEGHAHLEKVGIDGVKI---DVTH--L 412
 DB 858 ---KELTLEKKKAD-ILLSR--MLPKQVABRLKAG--QVPEPGFSPSVYFSDVYKFTI 909
 OY 413 LEMLCEDYGRVLDLAKVYKAMTKSINKHFKNGVGLASHECHDFLGTAEATSLGRVD 472
 DB 910 LMSKCSPP-QYVNLMDLYSNEDTITEQH---GV-----YKVESIGD 947
 OY 473 DEWCSTPSSDPNGTFWLGCHVHCANDSLMGNFTHPMDMFQSTHPCAAFRHASR--- 529
 DB 948 GYLCVSGPLTRNGY-----AHIKQIVDMSLKEMEY-----CKSEFIPLPRE 989
 OY 530 -----AISGGPIYVSDVSGKHNFDLLKLVLPDGSILNSEYALPTROCLFEDPLHGE 583
 DB 990 NMLRIGVNSGPC-VAGVVG-----LSMP-----RYCLEFDTVNTAS 1025
 OY 584 TM 585
 DB 1026 RM 1027

RESULT 12
 YCBL_BACSU STANDARD: PRT: 306 AA.
 AC P424242;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Hypothetical protein ycbJ.
 GN YCBL
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Azevedo V., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Kuster D., Bertolo M.G., Bessieres P., Bolotin A., Borcher S.,
 RA Bortis R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capiano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entlan K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
 RA Fritz C., Glaeser P., Goffeau A., Goldshly E.J., Galleron N.,
 RA Ghim S.Y., Glaeser P., Goffeau A., Goldshly E.J., Grandi G.,
 RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holseppel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Kohlnein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetle D., Portetle B., Rapoport G., Rey M., Reynolds S.,
 RA Presenc M., Rivolet C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Rieger M., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekinuchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin J., Taccori E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vasseroth A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Wetzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).

RN [2]
 RP SEQUENCE OF 98-306 FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95219079; PubMed=7704254;
 RA Ogawa K.-I., Akagawa E., Nakamura K., Yamane K.;
 RT "Determination of a 21348 bp nucleotide sequence around the 24
 RT degrees region of the Bacillus subtilis chromosome.";
 RL Microbiology 141:269-275(1995).
 CC -1 SIMILARITY: TO AMINOGLYCOSIDE PHOSPHOTRANSFERASES AND SPECIFICALLY,
 CC TO VIOMYCIN PHOSPHOTRANSFERASE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.1sb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC
 CC EMBL: 299105; CAB12046.1;
 CC EMBL: D30808; BAA06474.1;
 CC Subtilist: BG1165; ycbJ.
 CC
 CC DR Subtilist: BG1165; ycbJ.
 CC
 CC DR Subtilist: BG1165; ycbJ.
 CC
 CC KW Hypothetical protein, Transferase; Kinase; ATP-binding;
 CC Complete proteome.
 CC FT ACT_SITE 204
 CC SEQUENCE 306 AA; 34481 MW; 5EF0FD9DEC085977D CRC64;

Query Match 2.3%; Score 98; DB 1; Length 306;
 Best local Similarity 20.1%; Pred. No. 1.9;

Matches 54; Conservative 33; Mismatches 84; Indels 98; Gaps 11;

OY 120 VILEKSDSGRPVY-----FL-----LPIV 138
 DB 40 VYFANKDTGMPWLKRRPSDVERASAGITLAFILANULTADVPRMHTPELVYPL 99
 OY 139 ECFPRST-----OPGDDFV---DVCVESGSSKVVDSRSMLYLHAGD 181
 DB 100 KTRPAGIDLEOKQYVWNNHQPSPDFVTLADIAE-----LH-GTD 142
 OY 182 PFALVEAMKIVYTHGTFRLLEKTPPGIVDFGCT--WDAFYLTVPQVIEGVRHL 239
 DB 143 QISAGSGIEVIRPE--DRQMTADSMVDVKNKLGSTLMEWQKVVDDAYWPFSSSL 200
 OY 240 VDGCEPRLGLIDDGQSGISG-----HSDPTKEGMMQTVAGEQMPCRLKFOENYKF 292
 DB 201 IRHDLRPHILIDQNRVYGLDWTAKVADKDVLVQTINGEKETRLLEYDQ---- 257
 OY 293 RDVYVPRKATPRAGQKMAFIDELKGEF 321
 DB 258 -----AGGRIMAK-MOEHISEMQAAY 277

RESULT 13
 PIPA.DROME STANDARD: PRT: 1095 AA.
 AC P13217;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 1-phosphatidylinositol-4,5-bisphosphate phospholipase
 DE (EC 3.1.4.11) (phospholipase C).
 GN NORPA.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=8831074; PubMed=2457447;
 RA Bloomquist B.T., Shortridge R.D., Schneuwly S., Perdew M.H.,
 RA Montell C., Steiler H., Rubin G., Pak W.L.;
 RT "Isolation of a putative phospholipase C gene of Drosophila, norPA,

```

RT and its role in phototransduction."
RL Cell 54:723-733(1988).
CC -1- FUNCTION: THE PRODUCTION OF THE SECOND MESSENGER MOLECULES
CC DIACYLGLYCEROL (DAG) AND INOSITOL 1,4,5-TRISPHOSPHATE (IP3) IS
CC MEDIATED BY ACTIVATED PHOSPHATIDYLINOSITOL-SPECIFIC PHOSPHOLIPASE
CC ENZYMS. NORA IS INVOLVED IN PHOTOTRANSDUCTION.
CC -1- CATALYTIC ACTIVITY: 1-phosphatidyl-1D-myo-inositol 4,5-
CC bisphosphate + H(2)O = D-myo-inositol 1,4,5-trisphosphate +
CC diacylglycerol.
CC -1- SIMILARITY: DOMAINS X AND Y ARE CONSERVED IN DIFFERENT FORMS OF
CC PLC AND ARE ESSENTIAL FOR CATALYTIC ACTIVITY.
CC -1- SIMILARITY: CONTAINS 1 C2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: J03138: AAA28724.1; -
CC PIR: A31225: A31225.
CC HSSP: P10688: 10A5.
CC Flybase: FBgn0004625; noipa.
CC InterPro: IPR000008; C2.
CC InterPro: IPR001192; PI_PLC.
CC InterPro: IPR000909; PI_PLC_xdom.
CC InterPro: IPR001711; PI_PLC_X.
CC Pfam: PF00168; C2; 1.
CC Pfam: PF00387; PI_PLC-X; 1.
CC Pfam: PF00388; PI_PLC-X; 1.
CC PRINTS: PR00390; PHPLIPASEC.
CC ProDom: PD001202; PI_PLC_Y; 1.
CC SMART: SM00239; C2; 1.
CC SMART: SM00148; PLCX; 1.
CC SMART: SM00149; PLCX; 1.
CC PROSITE: PS50004; C2_DOMAIN_2; 1.
CC PROSITE: PS50007; PIPLC_X_DOMAIN; 1.
CC PROSITE: PS50008; PIPLC_Y_DOMAIN; 1.
CC HydroLase: Lipid degradation; Vision; Transducer.
CC FT DOMAIN 320 469 DOMAIN X.
CC FT 550 666
CC FT DOMAIN 673 771 C2 DOMAIN.
CC FT ACT_SITE 334 381 BY SIMILARITY.
CC FT ACT_SITE 381 BY SIMILARITY.
CC SEQUENCE 1095 AA; 124822 MW; 2D945BEF0ACB69B3 CRC64;

Query Match 2.38; Score 97.5; DB 1; Length 1095;
Best Local Similarity 19.48; Pred. No. 13;
Matches 109; Conservative 69; Mismatches 210; Indels 175; Gaps 26;

OY 144 TSIOPGDDVDV--CVESGSKYVDASFRSMY-LHAGD-----DPLFLVKAMKTVR 194
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 233 TSITKGGQDITSLFOIFQPMNDKORDRMANEILPLIEKRCTEITINDYELDEKKKNVQ 292
195 THLGTFR-LLEKTPRGIVDKFGWCTWDAFYL-----TYHPQ 230
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 293 MSIDGFRKRYLMSDNAPVFLDRD-----FYEMMDQPLAHYIINSSHTNYLSGRQIGK 346
OY 231 GVLGVRNHLVDGCGPRLVYIDGKWSIGHDSPTIREG-----MNOTVAGEQM 279
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 347 SSVEMRQTLTLAGC--RCVELDQWNGKGEDEEIVTHGAACTELLFKDCIOAIA---- 399
OY 280 PCRL--KFOENKFRDYVPKATGPRAGOGKKAFTDELKGEKTYEHYVNHALCGYV 337
400 DCAFVSEYVILSFENHCN-----RAQYKLAKTCDFFGDLLEKPL----- 443
DB 338 GGLRPQVPGV--PEARVIOPLVSPGLQMTMEDLAVDKIVLH-----K 377
||| ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB 444 -PDRPLDPGLRPPPCPKLKRIILKNNKRP--VEKVELLEMLKGLKTDDEEDAS 499
OY 378 VGLVP-----PEKAEMEYEGTLHAHLKEKVGIDGVKIDVYIHLLEMLCEDYGG 422

```

```

DB 500 AGKPEAAAAAPAPAEAAAAAEGAGCGGAEEAANAANYSGSTTNVHPMLSSMY-NYAQ 558
OY 423 RVDLAKYYVAMTSSIKRHKRGNVISMHC--NDPMFGTEILSIRGVDGDDPWCDDPS 480
DB 559 PIRF-QGFKAIETKINHNHNSFSAESAGMYLKOSIDFVYNNKRSRLYPKCTRADSS 617
OY 481 GDPNGTEFWLQGHVHCANDS-----LMMGNFTH-----PDMDMFGS 517
DB 618 NYMPQVFMNQCQGVVNLFPQSDPLPMQLNGKFEYNGCGVLLKPDPMRADKDFDFAD 677
OY 518 -----THPCAARHAASRAISGPT--YV-----SSSVCK-----H 545
DB 678 APYDGVIAACSVYVIAQFSLDSKRVGYEVDMFGLPSDTYKKEFRTLAVNGLNPNVY 737
OY 546 NFD--LLKLVLPDGSILRSEY 566
DB 738 NEDPFRKVVLPDLAVLREGVY 760

RESULT 14
HMCS_CRIGR
ID HMCS_CRIGR STANDARD; PRT; 520 AA.
AC P13704;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 4.1.3.5) (HMG-CoA
DE synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HMGCS1 OR HMCS.
OS Cricetus griseus (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10029;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=66140166; PubMed=2869035;
RA Gill G., Goldstein J.L., Slaughter C.A., Brown M.S.;
RT "Cytoplasmic 3-hydroxy-3-methylglutaryl coenzyme A synthase from the
RT hamster. I. Isolation and sequencing of a full-length cDNA."
RL J. Biol. Chem. 261:3710-3716(1986).
CC -1- FUNCTION: THIS ENZYME CONDENSES ACETYL-CoA WITH ACETOACETYL-CoA
CC TO FORM HMG-CoA, WHICH IS THE SUBSTRATE FOR HMG-CoA REDUCTASE.
CC -1- CATALYTIC ACTIVITY: (S)-3-hydroxy-3-methylglutaryl-CoA + CoA =
CC acetyl-CoA + H(2)O + acetoacetyl-CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-CoA PRIOR TO THE
CC SYNTHESIS OF STEROLS SUCH AS CHOLESTEROL AND ISOPRENOLIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE HMG-CoA SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L00334; AAA37076.1; -
CC EMBL: L00326; AAA37076.1; JOINED.
CC EMBL: L00327; AAA37076.1; JOINED.
CC EMBL: L00328; AAA37076.1; JOINED.
CC EMBL: L00329; AAA37076.1; JOINED.
CC EMBL: L00330; AAA37076.1; JOINED.
CC EMBL: L00331; AAA37076.1; JOINED.
CC EMBL: L00332; AAA37076.1; JOINED.
CC EMBL: L00333; AAA37076.1; JOINED.
CC PIR: A25332; A25332.
CC InterPro: IPR000590; HMG-CoA_synth.
CC Pfam: PF01154; HMG-CoA_synth; 1.
CC PROSITE: PS01226; HMG-CoA_SYNTHASE; 1.

```

KW Lyase; Cholesterol biosynthesis; Multigene family.
 FT ACT_SITE 129 129 POTENTIAL.
 SO SEQUENCE 520 AA; 57318 MW; 2DACAEE7DEE5D6BB CRC64;

Query Match 2.3%; Score 97; DB 1; Length 520;
 Best Local Similarity 18.8%; Pred. No. 5;

Matches 80; Conservative 63; Mismatches 115; Indels 166; Gaps 21;

```

OY 378 VGLVPE-----KAE-EMEGLAHLAEKVIDGKID-----VIHL 412
DB 17 VGIVALEIFPPSOYVDQALEKIDGVDACKYITIGLQAKMGFCTDIEDINSICLITYVQNL 76
OY 413 LE--MLCEDYGGKRVDLAKAYKAMTKSINKHF-----KNGVYASMEHCNDFMLGTEA 464
DB 77 MERNLSLSDYICIGLEVTETIDKSKSVKSNLMQLFEESGNTDIEGIDPTN--ACYGTTAA 135
OY 465 ISLGRVGDDFWCTDPSGDNGTFMLQGCCHVHCANDSLMKGNF---IHPDWMFQ--STH 519
DB 136 VF-----NAVMWIE-----SSWDGRVALVAGDIAIYATGNAR 169
OY 520 PCAAFHAAASRAISGPIYVSD-----SVGKHNFDLKLLVLPDGSILRSBYALPTRDCL 574
DB 170 PGCGVAAVALLIGPNAPLFDKRLGRTMHOAYDFK----PD---MLSEY-----213
OY 575 FEDPLHNGETMLKIMNLNKFVGIGAFNCOGGGWCRTRRNOCFSQYSKRVTSKTNPKDI 634
DB 214 ---PIVDGKLSIQ-----CYLSALDRCSYVRKKIRA-----242
OY 635 EWHSGENPISIEGVKTFALYLYQAKKLLSKPSQDIDLDPFEFLITVSPYTKLIOTS 694
DB 243 QW-----QKEGNDNDFTLNDFGF--MISHPYCKLVQKS 274
OY 695 LHFAPIGLVNMLNTSGAISOVDYDDLSVEIGKCGEKRFAVSKKPRACRIDGEDVGF 754
DB 275 L--ARMLNDFLN-----DONRDKNSIYSGLEAFGDVKL-----EDTYF 311
OY 755 KYDQDQ 760
DB 312 DRDVER 317

RESULT 15
HMGCS_RAT STANDARD: PRT: 520 AA.
AC P17425;
DT 01-AUG-1990 (rel. 15, Created)
DT 01-AUG-1990 (rel. 15, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Hydroxymethylglutaryl-CoA synthase, cytoplasmic (EC 4.1.3.5) (HMG-CoA synthase) (3-hydroxy-3-methylglutaryl coenzyme A synthase).
GN HMGCS1 OR HMGCS.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
ON NCBI_TaxID=10116;
RX MEDLINE=90301491; PubMed=1972979;
RP SEQUENCE FROM N.A.
RC STRAIN=Sprague-Dawley; TISSUE=Liver;
RA Ayte J., Gil-Gomez G., Hegardt F.G.;
RT "Nucleotide sequence of a rat liver cDNA encoding the cytosolic 3-hydroxy-3-methylglutaryl coenzyme A synthase."
RL Nucleic Acids Res. 18:3642-3642(1990).
CC -1- FUNCTION: THIS ENZYME CONDENSES ACETYL-CoA WITH ACETOACETYL-CoA TO FORM HMG-CoA, WHICH IS THE SUBSTRATE FOR HMG-CoA REDUCTASE.
CC -1- CATALYTIC ACTIVITY: (S)-3-hydroxy-3-methylglutaryl-CoA + CoA -> acetyl-CoA + H(2)O + acetoacetyl-CoA.
CC -1- PATHWAY: PRODUCTION OF MEVALONATE FROM HMG-CoA PRIOR TO THE SYNTHESIS OF STEROIDS SUCH AS CHOLESTEROL AND ISOPRENOIDS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE HMG-CoA SYNTHASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: X52625; CAA36852.1; -
 DR PIR: S12736; S12736.
 DR InterPro: IPR000590; HMG-CoA-synt.
 DR Pfam: PF01154; HMG-CoA-synt. 1.
 DR PROSITE: PS01226; HMG-CoA-synthase: 1.
 KW Lyase; Cholesterol biosynthesis; Multigene family.
 FT ACT_SITE 129 129 POTENTIAL.
 SO SEQUENCE 520 AA; 57433 MW; CB213A27B0C177CB CRC64;

Query Match 2.3%; Score 97; DB 1; Length 520;
 Best Local Similarity 18.8%; Pred. No. 5;

Matches 80; Conservative 62; Mismatches 117; Indels 166; Gaps 21;

```

OY 378 VGLVPE-----KAE-EMEGLAHLAEKVIDGKIDYIHLEMLCEDYGGKRVDLA 427
DB 17 VGIVALEIFPPSOYVDQALEKIDGVDACKYITIGLQAKMG-----FCTD--REDIN 66
OY 428 KAYKAMTKSINKHFKNGVYASMEHCNDFMLGTEAI-----SLGRVGDDFWCTD 478
DB 67 SICTLVYQKLMERNLSYDICIQLE-----VGTETIDKSKSVKSNLMQLFEESGNTD 119
OY 479 PSGDP-----NGTFMLQGCCHVHCANDSLMKGNF---IHPDWMFQ--THP 520
DB 120 IEGIDTTNACYGTTAAVFNAVMWIE-----SSWDGRVALVAGDIAIYASGNAR 170
OY 521 CAAAFHAAASRAISGPIYVSD-----SVGKHNFDLKLLVLPDGSILRSBYALPTRDCLF 575
DB 171 TCGCVAAVALLIGPNAPVIFDRGLRGTMHOAYDFK----PD---MLSEY-----213
OY 576 EDPLHNGETMLKIMNLNKFVGIGAFNCOGGGWCRTRRNOCFSQYSKRVTSKTNPKDI 635
DB 214 --PVVDGKLSIQ-----CYLSALDRCSYVRKKIRA-----Q 243
OY 636 WHSGENPISIEGVKTFALYLYQAKKLLSKPSQDIDLDPFEFLITVSPYTKLIOTS 695
DB 244 W-----QKEGNDNDFTLNDFGF--MISHPYCKLVQKSL 275
OY 696 HFAPIGLVNMNTSGAISOVDYDDLSVEIGKCGEKRFAVSKKPRACRIDGEDVGF 755
DB 276 --ARMLNDFLN-----DONRDKNSIYSGLEAFGDVKL-----EDTYF 312
OY 756 YDQDQ 760
DB 313 RDVER 317

```

Search completed: April 4, 2003, 11:11:49
 Job time : 36 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 4, 2003, 11:09:51 ; Search time 29 seconds
(without alignments)
795.433 Million cell updates/sec

Title: US-09-675-208-5

Perfect score: 4225

Sequence: 1 MAPSFKNNGSNVVSFDGLND.....QVPWPDSSSGISVIEYLF 784

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4225	100.0	784	4 US-08-846-234-5	Sequence 5, Appl1
2	181	4.3	30	4 US-08-846-234-1	Sequence 1, Appl1
3	109	2.6	1390	2 US-08-770-544-2	Sequence 2, Appl1
4	107.5	2.5	1584	4 US-09-251-645-6	Sequence 6, Appl1
5	107	2.5	487	4 US-09-004-838-101	Sequence 101, App
6	107	2.5	488	4 US-09-004-838-47	Sequence 47, Appl
7	107	2.5	1174	1 US-08-040-751-3	Sequence 3, Appl1
8	107	2.5	1174	1 US-08-291-368-2	Sequence 2, Appl1
9	107	2.5	1174	2 US-08-962-190-2	Sequence 2, Appl1
10	107	2.5	1174	5 PCT-US95-10310-2	Sequence 2, Appl1
11	107	2.5	1174	6 US-08-846-234-2	Sequence 2, Appl1
12	105	2.5	19	4 US-09-819-993-2	Sequence 2, Appl1
13	102	2.4	478	4 US-09-819-993-5	Sequence 5, Appl1
14	100	2.4	518	4 US-09-819-993-6	Sequence 6, Appl1
15	100	2.4	520	1 US-08-305-505-6	Sequence 8, Appl1
16	100	2.4	1276	1 US-08-717-515-8	Sequence 1, Appl1
17	98.5	2.3	770	1 US-08-525-654A-1	Sequence 3, Appl1
18	98.5	2.3	771	1 US-08-525-654A-3	Sequence 131, App
19	98	2.3	485	4 US-09-004-838-131	Sequence 48, Appl
20	97	2.3	472	4 US-09-004-838-48	Sequence 103, App
21	97	2.3	420	4 US-08-305-505-4	Sequence 4, Appl1
22	97	2.3	520	1 US-08-305-505-5	Sequence 4, Appl1
23	97	2.3	520	1 US-09-819-993-4	Sequence 28, Appl
24	96	2.3	772	4 US-09-134-078-28	Sequence 2, Appl1
25	96	2.3	712	1 US-08-587-889-2	Sequence 5, Appl1
26	95.5	2.3	712	1 US-08-980-060-5	
27	95.5	2.3	712	1 US-08-980-060-5	

28	95.5	2.3	712	4 US-09-307-185-5	Sequence 5, Appl1
29	95.5	2.3	712	5 PCT-US96-09193-2	Sequence 22, Appl1
30	95	2.2	600	4 US-09-388-743-22	Sequence 8, Appl1
31	94	2.2	750	4 US-09-364-970-8	Sequence 17, Appl
32	93.5	2.2	540	4 US-08-687-580B-7	Sequence 13, Appl
33	93.5	2.2	600	4 US-09-423-468A-13	Sequence 10, Appl
34	92.5	2.2	1423	4 US-08-810-712-10	Sequence 10, Appl
35	92	2.2	280	2 US-08-403-853-10	Sequence 21, Appl
36	92	2.2	649	4 US-09-066-047-5	Sequence 2, Appl1
37	92	2.2	820	4 US-09-313-677-21	Sequence 19, Appl
38	92	2.2	926	4 US-09-313-677-2	Sequence 17, Appl
39	92	2.2	933	4 US-09-313-677-19	Sequence 19, Appl
40	92	2.2	967	4 US-09-313-677-17	Sequence 2980, Ap
41	90.5	2.1	504	4 US-09-134-001C-2980	Sequence 15, Appl
42	90.5	2.1	592	4 US-09-423-468A-15	Sequence 108, App
43	89.5	2.1	1854	4 US-09-004-838-108	Sequence 15, Appl
44	89	2.1	559	4 US-09-242-690A-15	Sequence 6, Appl1
45	89	2.1	559	4 US-09-298-924-6	

ALIGNMENTS

```
RESULT 1
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
US-08-846-234-5

Query Match      100.0%; Score 4225; DB 4; Length 784;
Best Local Similarity 100.0%; Pred No. 0;
Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MAPSFKNNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS 60
      |||
Db      1 MAPSFKNNGSNVVSFDGLNDSSPFAIDGSDFTVNGHSFLSDVPENIVASPSPTSIDKS 60
      |||
QY      61 PVSVCFCVGFDASEPDSRHVSTGKLDKIDRFKSIFFRKVMTTHWVGRNGDLESEFOIV 120
      |||
```

```

Db      61 PVSVCNCFGEFSDSEPDSSRHVNSIGIKLIREMSIFREKYVMTTHWVGRNGDLESTQIV 120
QY      121 ILEKSDSGRPYFLLPIYEGPRTSIQPGDDFVDYCVESGSSKVVYDASFRSMLYLHAD 180
Db      121 ILEKSDSGRPYFLLPIYEGPRTSIQPGDDFVDYCVESGSSKVVYDASFRSMLYLHAD 180
QY      181 DPFALYKEMKIVRHLGLTFRLLEKTPPGIYDKFEGWCTWDAFYLTIVHOGVIEGRHLV 240
Db      181 DPFALYKEMKIVRHLGLTFRLLEKTPPGIYDKFEGWCTWDAFYLTIVHOGVIEGRHLV 240
QY      241 DGGCPGGLVLIIDGMSIHDSDPIKKEGNOTVAGEQPCRLKQOEYKFRDYVNPKA 300
Db      241 DGGCPGGLVLIIDGMSIHDSDPIKKEGNOTVAGEQPCRLKQOEYKFRDYVNPKA 300
QY      301 TGPRAQOKMKAFIDELKGEFTVEHVYVHALCGYWGSLRPQVGLPEARVIOPLVLSG 360
Db      301 TGPRAQOKMKAFIDELKGEFTVEHVYVHALCGYWGSLRPQVGLPEARVIOPLVLSG 360
QY      361 LQMTMEDLAVDKIVLHKVGLVPPERKAEMEYEGILHAHLEKVGIDGVKIDVILHEMLCEBY 420
Db      361 LQMTMEDLAVDKIVLHKVGLVPPERKAEMEYEGILHAHLEKVGIDGVKIDVILHEMLCEBY 420
QY      421 GGRVDAKAYYKAMTKRSINKHFKNGVITASMEHCNDFMLGTEALISLGVGDGDFMCTDS 480
Db      421 GGRVDAKAYYKAMTKRSINKHFKNGVITASMEHCNDFMLGTEALISLGVGDGDFMCTDS 480
QY      481 GDPNGTFWLOGCHMYVHCANDSLMNGFIHPDMDFQSTHPCAFHAASRAISGSPITYSD 540
Db      481 GDPNGTFWLOGCHMYVHCANDSLMNGFIHPDMDFQSTHPCAFHAASRAISGSPITYSD 540
QY      541 SVGKHNFDLKLVLPGDGLRSEYVALPTROCLEFDPHNETMLKINNLMKFTGVIA 600
Db      541 SVGKHNFDLKLVLPGDGLRSEYVALPTROCLEFDPHNETMLKINNLMKFTGVIA 600
QY      601 FPGCGGCMCRERTRNOCFOYSKRVTSKTPKDIEMHSGENPISIEGVTPLLYLOAKK 660
Db      601 FPGCGGCMCRERTRNOCFOYSKRVTSKTPKDIEMHSGENPISIEGVTPLLYLOAKK 660
QY      661 LILSRPSODLIALDPFEELITVSPVKLIQTSLHFAPIGLVNMMLTSGAIOQVYDD 720
Db      661 LILSRPSODLIALDPFEELITVSPVKLIQTSLHFAPIGLVNMMLTSGAIOQVYDD 720
QY      721 LSSVEIGVGCCEMVFASKPRACRIDEGVGFYDQOMVYVQVWPIIDSSGSIYI 780
Db      721 LSSVEIGVGCCEMVFASKPRACRIDEGVGFYDQOMVYVQVWPIIDSSGSIYI 780
QY      781 EYLF 784
Db      781 EYLF 784

```

RESULT 2
US-08-846-234-1
Sequence 1, Application US/08846234

GENERAL INFORMATION:
PATENT NO. 6166292
APPLICANT: OSUMI Chieko
APPLICANT: NOZAKI Jinshi
APPLICANT: KIDA Takao
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/846,234
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-413-3000
TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-846-234-1

Query Match 4.3%; Score 181; DB 4; Length 30;
Best Local Similarity 100.0%; Pred. No. 3.6e-12;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 215 FGCWCTWDAFYLYVHPQGVIEGVRLVYDGGC 244
Db 1 FGCWCTWDAFYLYVHPQGVIEGVRLVYDGGC 30

RESULT 3
US-08-770-544-2
Sequence 2, Application US/08770544

PATENT NO. 5907085
GENERAL INFORMATION:
APPLICANT: Gonsalves, Dennis
APPLICANT: Ling, Kai-Shu
TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND
THEIR USES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/770,544
FILING DATE:
CLASSIFICATION: A35
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60009008
FILING DATE: 21-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/621
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1390 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-770-544-2


```
QY 738 --ASKKPRACRIDGEDYGFKYDDQDMVVVQVWPWIDSSGGISVIEY 782
      | : | : | : | : | : | : | : | : | : | : | : | : |
Db 980 SYDAQRLVKTEDQDAQATFEYDNVGRLLITTTTKDTTSLSQLVTKIEY 102
```

RESULT 5

```

US-09-004-838-101
Sequence 101, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
CONFERRING PEST RESISTANCE IN PLANTS
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Elmhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-078810US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 101:
SEQUENCE CHARACTERISTICS:
LENGTH: 487 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: -
LOCATION: 1..487
OTHER INFORMATION: /note= "Kc2G deduced sequence"
US-09-004-838-101

```

Query Match	2.5%	Score 107	DB 4	Length 487
Best Local Similarity	20.1%	Pred. No. 0.035		
Matches 89	Conservative 63	Mismatches 167	Indels 124	Gaps 21

```

QY 237 RHLVDGGCPPELVILIDGMSIGHDS---DPITKEGNNQY---AGQMCRLLKPEYNT 250
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 61 RFEDAGCKNFELVILIDWQFVDEIDIGLSPLPKNGFNFVLLTSSRSHYCTLLMGAEAN 119
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 291 KFRDYVPKATGPAGOKGKMAFL-----DELKEFTEVHYVWMAHLCGYMGGLRQVP 345
      :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
Db 120 ---SILNIKVLKDEWGOSLEFROFKAGNAGDDDLDPAFN-----GIADSIASRCQ 164
      :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
QY 346 GLPEARIQVPLSLGLOMTMEDLAVDIYVLHKGVLPPEKAEMEYBEGLAHLEKVIDGY 405
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 165 GLPIAIFTIALSLTEGSKRSAMDVALSLRLENHKKIG---SEEVAREVF-----KISYDNL 214
      :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:
QY 406 KIDYIHLEMC---ED-----YGRDYL---AAAYTKAMT-----KSTN 439
      :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:  :|:

```

```

Db 215 QDEYKTSIFLLCALFEPDEFDIPTEBELVRYGMLKLFIEANTIRARNRRLTCTGERLRETN 274
OY 440 KHEKCN--GVYASMEHCNDF---MFLGTEAISLCRGVDDPWCTDPSGDPNGTMYLQCCIM 434
Db 275 LIFGSDIGGVCKMHDYVRDFVLHFFSEVQHASIVNHG-----NVSEWLEENHS 322
OY 495 VH-----CANDSLMWGNFIHFDMQMFQSTHPCAAFAASRAISGPIYVSQVSKH- 545
Db 323 IYSCRISLITLCKGMSQFPRKDLKFFNLSTLKLHM-----GDKSLT---FFPEFYCKME 371
OY 546 -----NEDLLKLLVLDG-----SLIRSEYVALPPIRDLFEDPLNGETM----- 585
Db 372 KVOYVISTYKMLYPLPLPSLEECSTNVRVLHLYCSLRMFDCSSIGLNLMEVLSFANSNIE 431
OY 586 ---LKIWLNKFTGVIGAFNCG 605
Db 432 WLPSTIGNLKLLR--LLDILTNCCK 453

```

RESULT 6

```

US-09-004-838-47
Sequence 47, Application US/09004838
Patent No. 6350933
GENERAL INFORMATION:
APPLICANT: Michelmore, Richard W.
APPLICANT: Shen, Kathy
APPLICANT: Meyers, Blake
TITLE OF INVENTION: Procedures and Materials for
TITLE OF INVENTION: Confering Pest Resistance in Plants
NUMBER OF SEQUENCES: 140
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/004,838
FILING DATE: 09-JAN-1998
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/781,734
FILING DATE: 10-JAN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Elnhorn, Gregory P.
REGISTRATION NUMBER: 38,440
REFERENCE/DOCKET NUMBER: 023070-078810US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 488 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: -
LOCATION: 1..488
OTHER INFORMATION: /note= "RUG2G protein"
US-09-004-838-47

```

Query Match	2.5%;	Score 107;	DB 4;	Length 488;
Best Local Similarity	20.1%;	Pred. No. 0.035;		
Matches 89;	Conservative 63;	Mismatches 167;	Indels 124;	Gaps 21;


```

Oy 237 RHLVGGCPGLVLLDDDMOSIGHDS---DPITKGNMOYV---AGEGMPRLKKEBENY 290
Db 62 RFEADGCKNKEVLIIIDDDWOFVDELEDIGSLPKNGVNFYLLTSSRDSHYCTLGAEN- 120
Oy 291 KFRDYVNPKRATGPRAQOKMKAFI-----DELKGEFKTEVHYVHNAICGYWGGLRPOV 345
Db 121 ---SLINKVLKDVGGSLFROFAKNAGDDDDLPAFN-----GIADSIASNCQ 165
Oy 346 GLPEKRVYQPVLSPOLQNTMEDLAVDKIVLKVGSLVPEPKAEEMEGJHAHLEKYGITGV 405
Db 166 GLPIAKITIALSLKGRKSAVDALSLRLENHKIG--SEEVREYV-----KISTYNL 215
Oy 406 KIDVHLEMLC---ED-----YGRVDV--AKAYUKAMT-----KSIK 439
Db 216 ODEVYKSTFLLCALFREDPDIPTBELVRYGWSGKLFIKAITIRKRNKLNLTERTLRRTN 275
Oy 440 KHFKN--GYIASMHCNDF--MFLGTEAISLGRVGDPMCTDPSGDPNGTFWLOGGIM 494
Db 276 LLFGSDDIGCVKMHNDVDFVLAHIFSEVONASIVNHG-----NVSEWLEENHS 323
Oy 495 VH-----CANDSLMKNFIHPMDMFOSTHPCAFHNASRAISGGRTIYVDSVGKH- 545
Db 324 IYSCRRISUTCKGMSQGFKDKLFNLSILKLMH-----GDKSLS--EPENYTGME 372
Oy 546 ----NFDLLKLVLPDG-----SILRSEYVALPTRDCLFEDPLHNGETM----- 585
Db 373 KYQVISTYCKLMLPLPSSLEECSTNVRYLHLHYSLRMDCSSIGNLIMVEVLFANSIE 432
Oy 586 ---LKIMNLNKFTGVIAGFNCOG 605
Db 433 WLPSTIGLKKLR-LLDLPNCKG 454

```

```

1      RESULT 7
2      US-08-040-751-3
3      : Sequence 3, Application US/08040751
4      : Patent No. 5407825
5      : GENERAL INFORMATION:
6      : APPLICANT: Payne, Jewel M
7      : APPLICANT: Sick, August J
8      : TITLE OF INVENTION: No. 5407825el Bacillus thuringiensis isolates
9      : TITLE OF INVENTION: active against Lepidopteran Pests and Genes Encoding No. 5407825el
10     : TITLE OF INVENTION: Lepidopteran-active Toxins
11     : NUMBER OF SEQUENCES: 4
12     : CORRESPONDENCE ADDRESS:
13     : ADDRESSEE: DAVID R. SALIMANCHIK
14     : STREET: 2421 N.W. 41st STREET, SUITE A-1
15     : CITY: GAINESVILLE
16     : STATE: FL
17     : COUNTRY: USA
18     : ZIP: 32606
19     :
20     : COMPUTER READABLE FORM:
21     : MEDIUM TYPE: Floppy disk
22     : COMPUTER: IBM PC compatible
23     : OPERATING SYSTEM: PC-DOS/MS-DOS
24     : SOFTWARE: PatentIn Release #1.0, Version #1.25
25     :
26     : CURRENT APPLICATION DATA:
27     : APPLICATION NUMBER: US/08/040.751
28     : FILING DATE: 19930329
29     : CLASSIFICATION: 435
30     :
31     : ATTORNEY/AGENT INFORMATION:
32     : NAME: SALIMANCHIK, DAVID R.
33     : REFERENCE/DOCKET NUMBER: MA39.C1.D3
34     : TELECOMMUNICATION INFORMATION:
35     : TELEPHONE: 904-375-8100
36     : TELEFAX: 904-372-5800
37     :
38     : TEXT:
39     : INFORMATION FOR SEQ ID NO: 3:
40     :
41     : SEQUENCE CHARACTERISTICS:
42     : LENGTH: 1174 amino acids
43     : TYPE: AMINO ACID
44     : STRANDEDNESS: single
45     : TOPOLOGY: linear

```

```

? MOLECULE TYPE: protein
?
? HYPOTHETICAL: YES
?
? ANTI-SENSE: NO
?
? ORIGINAL SOURCE:
? ORGANISM: Bacillus thuringiensis
?
? STRAIN: aizawai
?
? INDIVIDUAL ISOLATE: PS81A2
?
? IMMEDIATE SOURCE:
? LIBRARY: Lambdagem - 11 (tm) Library of August Slack
? CLONE: 81A2
?
US-08-040-751-3

```

Query Match	2.58;	Score 107;	DB 1;	Length 1174;
Best Local Similarity	19.68;	Pred. No. 0.17;		
Matches 11;	Conservative 82;	Mismatches 181;	Indels 192;	Gaps 29

```

0Y 336 YWGGJR-----POVGLDE-----ARVIOVLSGLQMTMEDLA-----VDK
Db 314 YWAGHRVTSHTFGSSQVLSSTPOYGITANAESKRTIADSTFGMLFRTISDPPFRSDN
0Y 373 I-----VLHKVGLVPEPEKAEMY--EGLHAHLEKVGIDGVR--IDVILHLEMLCEDYG
Db 374 IMPLTGINWVGVOGFIGQNNNEVLYRRGRGVDSIDELPIDGENSLVGYSHRL-----
0Y 422 GRVDLAKAYVYAMTKST-----NKRFKNGVYIASHECHNDP-MPLGTEAI-SLG
Db 426 SHVTLTRLSLYMTNTISLTPTFWTHHSATDRNIIPDVITQIPLVKSFSLTSGTSVWVGPG
0Y 469 RVGDGFMCTDPSGD-----PNGTFMLOGCHM
Db 466 FTGGIITRTNVNNGVNLMSLNFNMTSLDRVRYRYVYASQTMWVRVWVGSTTFDQGFPS
0Y 495 VHCANDSLMGNFIHPDMDFQSTHPCAAFHAAASRAISGCEPIYVSDSGVKNHFDLKKLV
Db 546 TWSAGSTLTSOSFRAEPEPVGIST-----SGSOTAG-ISISNPGROTFLDRIEF
0Y 555 LPDGSILMSEV---YALPTROCLFEDPLHNGETMUKI-----WNLNKFTGVGAFANCCGGG
Db 556 IPVDTEFEAEFDLEBAQKAVNSLFTS--SNQIEKTDYDHYHDOVSNLV--DCLSD
0Y 608 WCRETRRNQCES-OYSKRVTSKTN-----PKDIEWHSG-----EN
Db 650 FCLDEKRELSEKVKHAKRLSDERNLQDPNFRGINRQDRCMRGSTDITIOGCDVKEKN
0Y 642 PISIEGV--KTFALYLYQAKKLLSK-----PSQDLIDLALDF--EPELIT
Db 710 YVTLTGTEDECPYLYXG-KIDESKLYAKNRQYLRGYIEDSODLEIYLIAYNAKHETVN
0Y 684 VSPVTKLQTSLSHPEFIDLYMNLMTSGAISOVDYDDLS-----SWEIG
Db 768 VEGTGSIMPLVSE-SPICRCEPNR--CVPILENNPDLDCSRGEGEKAAHSHHPSLIDID
0Y 728 VKGCGEMRVFASKKRPACRIDGEDVG 753
Db 825 V-GCTDLO-----EDLG 835

RESULT 8
US-08-291-368-2
; Sequence 2, Application US/08291368
; Patent No. 5686069
; GENERAL INFORMATION:
; APPLICANT: Payne, Jewel M.
; TITLE OF INVENTION: No. 5686069el Bacillus thuringiensis Isolates
; TITLE OF INVENTION: Active Against Lepidopteran Pests
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanichik & Saliwanichik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: FL
; COUNTRY: US

```



```

Db 314 YWAGHRTVSHFTGSSQVSSPOYGITANAEDSRITAPSTFPGNLFFYRTLSDPFFRRSDN 373
Qy 373 I-----VLHKVGLVPEPKAEEMT--EGLHAHLEKVGIDGK--IDYIHLEMLCEDYG 421
Db 374 IMPTLGIVNVOGVGFIOQNNGEVLYRRRGTYDSDELPIDGENSLVGSYSHRL----- 425
Qy 422 GRVDLAKAYVKAMTKSI-----NKHFGKNGVIAEMHCNDF-MFLGTEAI-SLG 468
Db 426 SHVTLTRSLYNTNITSLPTFWTHHSATDRNLIYDPVITQIPLVKSFLSTGTSVVRGPG 485
Qy 469 RVGDDEFWCTDPSGD-----PNCFFMLQGGCHM 494
Db 486 FTGGDIIRTNVNGVNLMSLNFSTSLQRYRVRYAASQTMWVRVNGGSTTFDQGFPS 545
Qy 495 VHCANDSLMGNFIHPDMDFOSTHPCAFHAASRAISGPIYVSDYSGKHNFDLKLTV 554
Db 546 TMSANGSLTSSQSFRAEPPVGIST-----SGSQTAG-ISISNNGRQTFHLDRIEF 595
Qy 555 LPDGSILRSEY--YALPTRDCLFEDPLHNGETMLKI---WNLKKFTGVIGAFNCGGG 607
Db 596 IPVDATFEAEYDLERAQAANSFLTSS--SNOIELKTDVTDYHIDQVSNLV--DCLSD 649
Qy 608 WCRETNRNOCES-QYSKRVTSKTN-----PKDIEMHSG-----EN 641
Db 650 FCLDEKRELSEKVKHAKRLSDERNLLODPNFRGINRQPDGRGSGTDITIQGGDVFKEN 709
Qy 642 PISIEGV--KTFALVLYQAKKLILSK-----PSQDDIALDP--EFELIT 683
Db 710 YVTLPGTDECYPYLYO--KIDESKLAAYNRQYLGRTIEDSODELITLIRYNAKHETVN 767
Qy 684 VSPVTKLIQTSLHFAPIGLVNMLNTSGAISOVDYDDLS-----SVEIG 727
Db 768 VPGTSLPLSYE--SPIRGCEPNR--CVPHLEMPDLDCSCRODEKCAHSHHPSLID 824
Qy 728 VKGCEMRVFAASKRPACRIDGEDVG 753
Db 825 V-GCTDLO-----EDLG 835

```

```

RESULT 10
: Sequence 2, Application PC/7US9510310
: GENERAL INFORMATION:
: APPLICANT: MYCOGEN CORPORATION 5501 Oberlin Drive
: APPLICANT: STREET ADDRESS:
: APPLICANT: CITY: San Diego
: APPLICANT: STATE/PROVINCE: California
: APPLICANT: COUNTRY: US
: APPLICANT: POSTAL CODE/ZIP: 92121
: APPLICANT: PHONE NUMBER: (619) 453-8030
: APPLICANT: FAX NUMBER: (619) 453-6991
: TITLE OF INVENTION: Protein Toxins Active Against Lepidopteran Pests
: NUMBER OF SEQUENCES: 27
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Saliwanichik & Saliwanichik
: STREET: 2421 N.W. 41st Street, Suite A-1
: CITY: Gainesville
: STATE: FL
: COUNTRY: US
: ZIP: 32606
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/10310
: FILING DATE:
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/597,607
: FILING DATE: 15-OCT-90
: CLASSIFICATION:

```

```

: ATTORNEY/AGENT INFORMATION:
: NAME: Saliwanichik, David R.
: REGISTRATION NUMBER: 31,794
: REFERENCE/DOCKET NUMBER: MA50.C1
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (904)375-8100
: TELEFAX: (904)372-5800
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1174 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHETICAL: YES
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Bacillus thuringiensis
: STRAIN: aizawai
: INDIVIDUAL ISOLATE: PS81A2
: IMMEDIATE SOURCE:
: LIBRARY: LambdaGen - 11 (tm) Library of August Sick
: CLONE: 81A2
: PCT-US95-10310-2

```

```

Query Match 2.5%; Score 107; DB 5; Length 1174;
Best Local Similarity 19.6%; Pred. No. 0.17;
Matches 111; Conservative 82; Mismatches 181; Indels 192; Gaps 29;

```

```

Qy 336 YWGLR-----POVGLPE-----ARVIOPLYSPGLQMTMEDLA-----YDK 372
Db 314 YWAGHRTVSHFTGSSQVSSPOYGITANAEDSRITAPSTFPGNLFFYRTLSDPFFRRSDN 373
Qy 373 I-----VLHKVGLVPEPKAEEMT--EGLHAHLEKVGIDGK--IDYIHLEMLCEDYG 421
Db 374 IMPTLGIVNVOGVGFIOQNNGEVLYRRRGTYDSDELPIDGENSLVGSYSHRL----- 425
Qy 422 GRVDLAKAYVKAMTKSI-----NKHFGKNGVIAEMHCNDF-MFLGTEAI-SLG 468
Db 426 SHVTLTRSLYNTNITSLPTFWTHHSATDRNLIYDPVITQIPLVKSFLSTGTSVVRGPG 485
Qy 469 RVGDDEFWCTDPSGD-----PNCFFMLQGGCHM 494
Db 486 FTGGDIIRTNVNGVNLMSLNFSTSLQRYRVRYAASQTMWVRVNGGSTTFDQGFPS 545
Qy 495 VHCANDSLMGNFIHPDMDFOSTHPCAFHAASRAISGPIYVSDYSGKHNFDLKLTV 554
Db 546 TMSANGSLTSSQSFRAEPPVGIST-----SGSQTAG-ISISNNGRQTFHLDRIEF 595
Qy 555 LPDGSILRSEY--YALPTRDCLFEDPLHNGETMLKI---WNLKKFTGVIGAFNCGGG 607
Db 596 IPVDATFEAEYDLERAQAANSFLTSS--SNOIELKTDVTDYHIDQVSNLV--DCLSD 649
Qy 608 WCRETNRNOCES-QYSKRVTSKTN-----PKDIEMHSG-----EN 641
Db 650 FCLDEKRELSEKVKHAKRLSDERNLLODPNFRGINRQPDGRGSGTDITIQGGDVFKEN 709
Qy 642 PISIEGV--KTFALVLYQAKKLILSK-----PSQDDIALDP--EFELIT 683
Db 710 YVTLPGTDECYPYLYO--KIDESKLAAYNRQYLGRTIEDSODELITLIRYNAKHETVN 767
Qy 684 VSPVTKLIQTSLHFAPIGLVNMLNTSGAISOVDYDDLS-----SVEIG 727
Db 768 VPGTSLPLSYE--SPIRGCEPNR--CVPHLEMPDLDCSCRODEKCAHSHHPSLID 824
Qy 728 VKGCEMRVFAASKRPACRIDGEDVG 753
Db 825 V-GCTDLO-----EDLG 835

```

```

RESULT 11
5164180-4
: Patent No. 5164180

```


Db 157 -----QHAIDFYK---PD---MLSEY-----PIVDGKL 178
QY 585 MLKIMNLKFTGIVGAFNCGGCMGRTERRNOCFSQYSKRVTSKTNPKDIEMHSGENPIS 644
Db 179 SIO-----CYLSALDCCYSVYCKKHA-----QW----- 202
QY 645 IEGVTFALYVQAKKLISKPSODDIALDPFEFLITVSPVKLTQSTLHFAPIGLVN 704
Db 203 -----QKGNOKDFTLNDGFG-MIFHSPYCKLVQKSL--AKMLND 240
QY 705 MLNTSGAIOVDYDDLSVEIGVKGCEMRVFASKRPACRIDEDVGFKYDQD 760
Db 241 FLN-----DQNRDKNSIYSGLEAFGDKL-----EDTYFDRDVER 275

RESULT 14
US-09-819-993-5
Sequence 5, Application US/09819993
Patent No. 6436692

GENERAL INFORMATION:
APPLICANT: GONG, Fangcheng et al.
TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
FILE REFERENCE: CLO01195
CURRENT APPLICATION NUMBER: US/09/819,993
CURRENT FILING DATE: 2001-03-29
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 518
TYPE: PRT
ORGANISM: Human
US-09-819-993-5

Query Match 2.4%; Score 100; DB 4; Length 518;
Best Local Similarity 19.0%; Pred. No. 0.21;
Matches 82; Conservative 59; Mismatches 111; Indels 180; Gaps 20;

QY 378 VGLVPE-----KAE-EMVEGLHAHLEKYDGVKID-----VHIL 412
Db 17 VGIVALEIYFSPQYVDQALEKYDGVDAKTYTIGLOAKMGCTDREDINSICMTVQNL 76
QY 413 LEM--LCEDYGRVDLAKAYKAMTKSINKHF-----KNGVIAHMEHNDPMFLGTEA 464
Db 77 MERNNLSTYDCIGRLEVGETITIDKSKSVKTNLMOLFEEBSGNTDIEGIDTTN--ACYGCTAA 135
QY 465 ISLGRVGDGFWCTDPSGDPNGTFMLQGCIMVHCANDSLMGNFIHPDMDFQSTHPCAA 524
Db 136 VF-----NAVMWIE-----SSWDGRY-----ALVAGDIAY 163
QY 525 HAASRAISGPIYVSDSVG-----KHNFDLKLKLVLPDGSILRSEYVAL 568
Db 164 ATGNARPTGCGAVALLIGPNAPLIFERGLRGTNHQHAIDFYK---PD---MLSEY--- 213
QY 569 PTRDCLFEDPLNGETMLKIMLNKFTGIVGAFNCGGCMGRTERRNOCFSQYSKRVTSK 628
Db 214 -----PIVDGKLSIO-----CYLSALDRCYSVYCKKHA- 242
QY 629 TNPKDIEMHSGENPISIEGVTFALYVQAKKLISKPSODDIALDPFEFLITVSPVT 688
Db 243 -----QW-----QKANDNDFTLNDGFG-MIFHSPYC 268
QY 689 KLIOTSLHAPRIGLVNMLNTSGAIOVDYDDLSVEITVKCGCEMRVFASKRPACRID 748
Db 269 KLVQKSL--AKMLNDFLN-----DQNRDKNSIYSGLEAFGDKL----- 306
QY 749 GEDVGFKYDQD 760
Db 307 -EDTYFDRDVER 317

RESULT 15
US-08-305-505-6
Sequence 6, Application US/08305505
Patent No. 5668001

GENERAL INFORMATION:
APPLICANT: Maziorko, Henry M.
TITLE OF INVENTION: 3-HYDROXY-3-METHYLGUTARYL-COA
TITLE OF INVENTION: SYNTHASE PREPARATION WITH IMPROVED
TITLE OF INVENTION: STABILITY
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quatles & Brady
STREET: 411 East Wisconsin Avenue
CITY: Milwaukee
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/305,505
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/072,040
FILING DATE: 02 JUNE 1993
ATTORNEY/AGENT INFORMATION:
NAME: Baker, Jean C.
REGISTRATION NUMBER: 35,433
REFERENCE/DOCKET NUMBER: 65-053-9083-9
TELECOMMUNICATION INFORMATION:
TELEPHONE: (414) 277-5709
TELEFAX: (414) 277-5591
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 520 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-305-505-6

Query Match 2.4%; Score 100; DB 1; Length 520;
Best Local Similarity 19.0%; Pred. No. 0.22;
Matches 82; Conservative 59; Mismatches 111; Indels 180; Gaps 20;

QY 378 VGLVPE-----KAE-EMVEGLHAHLEKYDGVKID-----VHIL 412
Db 17 VGIVALEIYFSPQYVDQALEKYDGVDAKTYTIGLOAKMGCTDREDINSICMTVQNL 76
QY 413 LEM--LCEDYGRVDLAKAYKAMTKSINKHF-----KNGVIAHMEHNDPMFLGTEA 464
Db 77 MERNNLSTYDCIGRLEVGETITIDKSKSVKTNLMOLFEEBSGNTDIEGIDTTN--ACYGCTAA 135
QY 465 ISLGRVGDGFWCTDPSGDPNGTFMLQGCIMVHCANDSLMGNFIHPDMDFQSTHPCAA 524
Db 136 VF-----NAVMWIE-----SSWDGRY-----ALVAGDIAY 163
QY 525 HAASRAISGPIYVSDSVG-----KHNFDLKLKLVLPDGSILRSEYVAL 568
Db 164 ATGNARPTGCGAVALLIGPNAPLIFERGLRGTNHQHAIDFYK---PD---MLSEY--- 213
QY 569 PTRDCLFEDPLNGETMLKIMLNKFTGIVGAFNCGGCMGRTERRNOCFSQYSKRVTSK 628
Db 214 -----PIVDGKLSIO-----CYLSALDRCYSVYCKKHA- 242
QY 629 TNPKDIEMHSGENPISIEGVTFALYVQAKKLISKPSODDIALDPFEFLITVSPVT 688
Db 243 -----QW-----QKANDNDFTLNDGFG-MIFHSPYC 268

Fri Apr 4 12:31:30 2003

us-09-675-208-5.raii

Page 10

Oy	689	KLIGSLFPAFIGVNMNTSGALQSVDDYDSDSSVEIGKGGCGEMVFASKMPACRID	748
Dd	269	KLVGKSLL--AAMLDNDFLN-----DQNDRKNSTYSGLKAAGDYVL-----	306
Oy	749	GEDVGCFKYDDO 760	
Dd	307	-EDITYRDNVER 317	

Search completed: April 4, 2003, 11:14:05
Job time : 34 secs


```
OY 301 TGPRAQKMKAFIDE-----LKGEFTEVHYVW-HAL-CGYNGULRPQVPLPEAR 351
DB 243 ---NMSSSESDSCDSTSETCVPLEGEKSAEKKRVRFOHLSCLINDMRKLLPEF----- 295
OY 352 VIQPVSPGLQMTME---DLAVDKIVLHKV-----GLVPEKAEMEYEGHLHALEK 399
DB 296 VROFSIDRCCEMTPKTPGDLAWN--FLMKYQALDSTARSDILRPEVAGEEENELPAGITEK 333
OY 400 VGIDGKVIDVHILLE-----MLCEDYGGRVDLAKAYYKAM-----TKSINKHFKGNG 446
DB 354 LGIGDPQ--TIHPLDVLCACMLCADSSLRREVSNMYQCOFALPRLPDAENKNKLLMWG 411
OY 447 VIASMEHCN-----DPMFLGTEAISLGRVGDDEFWCTDPSGDP 483
DB 412 AMKDLQPSQOSSGGLRETFELGLTKMPVISFVRLGRCSFSKSRIVNTL--LSSSQOKP 470
OY 484 NGTFWLGCH--MVHCANDSLMGNFIHPDMDPOSTH---PCAFAHNR----- 529
DB 471 YPIFLHQDLSVPVLPQISGLVEVTWCSPDKLLKESPHAFQKPVAVANLRGDLSEFWIQ 530
OY 530 ---AISGPIYVSDVSGKHNFEDLLKLVLPDGSILRSEYALPTRDCLFEDPLHNGET 584
DB 531 FGLVEVSSGLFEFTDCLGEKENDL--MFLGSDTIERCFILSPQ----- 574
OY 585 MKTIMNLNKTGVIYAFNCOGGWCRTERRNOCFSQYSKRWYSKTPKDIEM--HSGEN 641
DB 575 -----AKESBAQIFORILKLPQSL-----LFWEAEAGDR 606
OY 642 PISIEGVTFEALYQAKLLISKP-----SODLIALDPEFEEL---ITVSP 686
DB 607 RKTMEA-----LQALDQVWSSPLRCVSLRMAASLARLGIQVDO--DEVTQDQIVSP 658
OY 687 VT 688
DB 659 TT 660

RESULT 2
US-10-193-295-2
; Sequence 2, Application US/10193295
; Patent No. US20020173018A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; FILE REFERENCE: CLO01195DIY
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 478
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-2

Query Match
Best Local Similarity 18.5%; Pred. No. 0.63; Length 478;
Matches 77; Conservative 54; Mismatches 95; Indels 190; Gaps 19;

OY 378 VGLVPE-----KAE-EMYEGLHALEKVIDGVRID-----VIHL 412
DB 17 VGLVLAIEIFPSQYVDQALEKDYDACKYITIGLOAAMGCTDREDINSICMTYVONI 76
OY 413 LEM--LCEDYGGRVDLAKAYYKAMTKSINKHF-----KNGVYIASMEHCNDFMLGTEA 464
DB 77 MERNNLSTYDICIARLEVEGTETIIDKSKSVKTNLMQLFEEGNDIDIEDITTN--ACYGTAA 135
OY 465 ISLGRVGDDEFWCTDPSGDPNGTFWLGCHMVHCANDSLMGNFIHPDMDPOSTHCAAF 524
```

```
DB 136 VF-----NAVNWIES-----SSWDGLRGTHM----- 156
OY 525 HAASRAISGPIYVSDVSGKHNFEDLLKLVLPDGSILRSEYALPTRDCLFEDPLHNGET 584
DB 157 -----QHAYDFK---PD---MLSTY-----PIVDGKL 178
OY 585 MKTIMNLNKTGVIYAFNCOGGWCRTERRNOCFSQYSKRWYSKTPKDIEMHSGENPIS 644
DB 179 SIQ-----CYLSALDRCSYCYCKIHA-----QW----- 202
OY 645 IESEVTFEALYQAKLLISKPSODLIALDPEFEELITVSPYTKLIQTSILHRAPIGLVN 704
DB 203 -----QKEGNDKPFITNDQGF--MIFHSPYCKLVQXSL--ARMLND 240
OY 705 MLNTGAISQVDYDDLLSVEIGVCKGCEMRVFAKKPRACRIDGSDVGFQYDQDQ 760
DB 241 FLN-----DGNRRKNSTIYSGLEAFAGVKT-----EDTYFDRDYEK 275

RESULT 3
US-10-001-851-27
; Sequence 27, Application US/10001851
; Patent No. US20020115628A1
; GENERAL INFORMATION:
; APPLICANT: MEYERS, Rachel A.
; APPLICANT: WILLIAMSON, Mark
; TITLE OF INVENTION: 47169 and 33935, No. US20020115628A1 Human Glycosyl Transfer
; TITLE OF INVENTION: Uses Thereof
; FILE REFERENCE: 10147-5601
; CURRENT APPLICATION NUMBER: US/10/001,851
; CURRENT FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: US 60/249,939
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 626
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-001-851-27

Query Match
Best Local Similarity 21.3%; Pred. No. 1.1; Length 626;
Matches 77; Conservative 52; Mismatches 118; Indels 115; Gaps 20;

OY 220 WDAFYLVTHPGVIEGV-RHLVDGCGPGLVLIIDGWSIGHSDPITREGMNQTVAGEO 278
DB 188 WSVLRTVH--SVLEKTPHILE-----EVLVDD--FSDMDHTRRL--EYMSQ--FGKV 237
OY 279 MPCRLIKFOENKFRDYVNKRATG-----RAGQKMAFIDELKGERFY----- 324
DB 238 KILRMEKREGILIRARLGAAVATGEVLYTLDSHCCEMGMEBLDRIRKRDPTTVCPVI 297
OY 325 ---EHVYVWH-----ALCGYWGGLRPQVRLPE-----ARYQVPLSPQLQMTMED 367
DB 298 DAVIDMTFEYHNSKAYFTSVGGFWGLQFNWHSIPBRDRKNRTPTIDPASP---TMAG 353
OY 368 --LAVDKIVLHKV-----LVPEKAEMEYEGHLAHU 397
DB 354 GLEFIDKEFEKIGTIDPDPGDTMGGENLISFKIMCGGLEIVPESHGHVFRKRSPTK 413
OY 398 EKVAGIDVKRIVDVIHILLEMLCEDYGGRVDLAKAYYKAMTKSINKHFGNGVIASMEH--- 453
DB 414 WRGVAVVLKNSIRLAELVWLDY-----KTY--YERINNQLGDFDGISSRKKLRD 463
OY 454 ---CNDP-----MELGTEAISLGRVGD---FWCTD-----PSGDNGTFTWLOG 491
DB 464 LOCKSRKWTLDNIYPLFLFVPGESVAVAGEVNSAVOPARCIDCWGGRHKNRPVGYT--Q 520
OY 492 CH 493
DB 521 CH 522
```



```
RESULT 4
US-10-193-295-5
; Sequence 5, Application US/10193295
; Patent No. US20020173018A1
; GENERAL INFORMATION:
; APPLICANT: GONG, Fangcheng et al.
; TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
; TITLE OF INVENTION: THEROP
; FILE REFERENCE: CLO01195D1V
; CURRENT APPLICATION NUMBER: US/10/193,295
; CURRENT FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 08/819,993
; PRIOR FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Human
US-10-193-295-5

Query Match          2.4%; Score 100; DB 9; Length 518;
Best Local Similarity 19.0%; Pred. No. 1.1;
Matches 82; Conservative 59; Mismatches 111; Indels 180; Gaps 20;

OY 378 VGLVPE-----KAE-EMEGHLEKVGIDGVKID-----VHIL 412
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 17 VGLVLEIYFPOSYQVDALEKRYDGDACKYITIGLQAKMGFCTDREDINSICMTVQNL 76

OY 413 LEM--LCEDYGRVDLAKAYKAMTKSINKHP-----KNGVIASMEHCNDPMEFGTEA 464
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 77 MRRNNLSYVICIRLEVGFTITIDKSKSVKTNLMQLFEESGNDIDISIDITN-ACIGGTAA 135

OY 465 ISLGRVDDFWCTDPSGDENGTFWLOGCHMVHCANDSLMGNFIHPDWMDFOSTHPCAAF 524
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 136 VF-----NAVNMIE-----SSWDGRY-----ALVAGDIAYV 163

OY 525 HAASRAISGPGIYVDSVG-----KHNDLKLKLVLPDGSILRSYVAL 568
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 164 ATGNARPTGCGAVALLGPNAPLIERGLRGTTHQHAIDFYK---P---MLSEY--- 213

OY 569 PTRDCLFEDPLHNGETMLKIMLNKFTGIVIGAFNCGGCMGCRETRNOCFSQYSKRVTSK 628
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 214 -----PIYDGRKLSIO-----CFLSALDRCTSYVCKIHA- 242

OY 629 TNPKDIEMHSGENPISIEGVKTFALYLYQAKKLILSKPSQDLIDALDPEFELITVSPVT 688
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 243 -----QW-----QKEANDNDFTLNDRGF-MIFHSPLYC 268

OY 689 KLIQTSIHAPFGLVNMNTSGAIOVDYDDDLSSVEIYKCGGKRVASAKPRACRID 748
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 269 KIVQKSL-ARMLNLDLFLN-----DQNRDKNSIYSLGKAFGDKL- 306

OY 749 GEDVGFKYDQDQ 760
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 307 -EDTYFDRDYEK 317

RESULT 5
US-10-095-718-4
; Sequence 4, Application US/10095718
; Patent No. US20020131956A1
; GENERAL INFORMATION:
; APPLICANT: Walsh, Christopher
; APPLICANT: Chao, Hengjun
; APPLICANT: Bursstein, Haim
; APPLICANT: Lynch, Carmel
; APPLICANT: Stepan, Tony
; APPLICANT: Munson, Keith
; TITLE OF INVENTION: Adeno-Associated Virus Vectors Encoding Factor VIII and
; TITLE OF INVENTION: Methods of Using the Same
```

```
; FILE REFERENCE: 350522/204375
; CURRENT APPLICATION NUMBER: US/10/095,718
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/689,430
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/158,780
; PRIOR FILING DATE: 1999-10-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1431
; TYPE: PRT
; ORGANISM: canine B-domain deleted factor VIII
US-10-095-718-4

Query Match          2.3%; Score 98; DB 12; Length 1431;
Best Local Similarity 19.0%; Pred. No. 8.2; Indels 316; Gaps 33;
Matches 133; Conservative 65; Mismatches 186; Indels 316; Gaps 33;

OY 11 NVVSFDGLNDSSPFAIDSDPTVNGH-----SFLSDVENIVASPSPTSIDKS----- 60
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 369 DVSVD--DDSSPFLQIRS--VAKKHPTWVHYIAEEDMDYAPSGFTPDNRSHKILY 424

OY 61 ---PVSVGCFVGFAPDASEPDSRHVVSICKLDIRPMSIFRFKVMWTTTHVGNNGDL--- 113
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 425 LNNGPQRIQ-----KKYKVRFAA-VTDEFPTREALIOYESGILGPL 465

OY 114 ---ESTQIVILEKSDSGRPYFLLPIYVGSPRTSIQPDDEDFVDYCVSSGSKYVDASF 170
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 466 LYGEVGDTLILIFKNQASRP-----NIYPGIYVVP----- 498

OY 171 RSMVLVHAGDDPFALVKEAMKIVRTHLGTRFLLEKTPPGIYDKFGQCTW-----DA 222
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 499 -----LHTGRLPKGV-----KILKDPIL-----PGEIYFKYKITYVEDGPRKSP 539

OY 223 FYLVNHPQGVIEGVRHLVDGCRPGLVLIDGMQSIGHSDPITREKGMQTVAGEQMPCR 282
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 540 RCTTRYSSFINLERDLASGLIGLPLICYKESVDQG-----NQMSDKRNVIL 588

OY 283 LKFOENYKFRDYVNPKATGPRAQGMKAFIDELKGEFKTYEYVYVHIALCGYMGGLRP 342
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 589 FSVFEDENRSM--YLQ-----ENMORF----- 607

OY 343 QVPGLPPEARVIOVPSGLQMTMEDLAVDKIVLHKVGLVPRPEKAEEMYGLAHLEKVG 402
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 608 ---LPNADYVP-HDPERQLSN-----IMH-----SI 630

OY 403 DGVKIDVILHLEMLCE-----DYGGRVDLAKAYKAMTKSINKH-----FK 443
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 631 NGYVFDNLQLSVCLHEVAWYILTSVGAQTFDLSVFPSGYT---FKHKMYEDTLTLFPSS 687

OY 444 GNGVIASME-----HCNDPMELGTEAT---SLGAVGDDFW----- 475
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 688 GETVFSMSENPGLAWLGLCHNSDFRNRGMETALLKVSNCNRNIDYEDYEDIPPTLNEN 747

OY 476 -----CTDPSG-----DPNGTFMLQGC-----HMVHCANDSLAM 504
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 748 NVIKPRFSQNSRHSTKQOLKMKREDFDYGDVENQGLRSFOKTRTYFTAVERLNL- 806

OY 505 GNFIFHPDMDFOSTHPCAFHAASRAISGPGIYVDSVGKHNHFDLKLVL---LPDGSIL 561
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 807 -----DYGMSRSPH-----ILNRNROSG-----DYQGRKVVYQDEFPTDGS-- 841

OY 562 RSEYIALPTRDCLFEDPLHNGETMLKIMLNKFTGIVIGAF 601
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 842 -----FTQPLRYGE-----LNEHLGILGPY 861

RESULT 6
US-09-978-522-1
; Sequence 1, Application US/09978522
; Publication No. US2003003627A1
; GENERAL INFORMATION:
```


FILING DATE: 13-AUG-1998
APPLICATION NUMBER: 08/949,026
FILING DATE: 10-OCT-1997
APPLICATION NUMBER: 60/056,916
FILING DATE: 06-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Hall, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/024002
TELEPHONE: 858/677-1456
TELEFAX: 858/677-1465
INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 772 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
SEQUENCE DESCRIPTION: SEQ ID NO: 28:
US-10-121-032-28

Query Match 2.3%; Score 96; DB 9; Length 772;
Best Local Similarity 20.6%; Pred. No. 4.8;
Matches 138; Conservative 64; Mismatches 166; Indels 302; Gaps 42;

QY 97 FKVM-----WTHWYGRNGDLESETOIYILEKSDG-----RPVFLPIVE--G 140
DB 165 FVWSPVSKWVAVLLEKNEDETE-PRVYVNMEEKGVWEAVEGDLDGCVLYOLENYG 223
QY 141 PRTSIOPDDDFVDVVCVSSGSSKVVDASFRLMLYHAGDDPFALV----- 187
DB 224 KRTTYDP-----YSKAV-----YANNQSAVVNLTARTNESEGENDR 260
QY 188 -----EAMKIVRTHLGTFRLLLEKTPPGIVDKGWCMTDAFYL-----TVHPOGVIE 234
DB 261 GKRIEGEYDAIYIEHIAIDITGLENS--GVKNK-----GLYIGLIEENTKPGCVTT 310
QY 235 GVRHLVDGCGPGLVLIDGCMQSIGHSDSPITKESNMOTVAGQOMCRLLKROENKFKFD 294
DB 311 GLSHLVELG-----VTH-----VHILPFFDFY----- 332
QY 295 YVNPKATGPRAGOKMKAFIDELKEGFTEHVYVNHALCGYWGSLRPQVGLPEARVIO 354
DB 333 -----TG-----DELDRDF--EKYVNM-----GYDPLFVMPREGRYST 363
QY 355 PVLSPLOMTMEDLAVDKIVLHKVGLVPEPKAEEMTEGLHAHLEKVGIGVKIDVTH--- 411
DB 364 DPKNPHTRI-----REVKEVYKALHKH---GI-GVIMDWVFPHT 398
QY 412 --LLEMLCED-----YGRVNDLAKAVYKAMTSSINKHFKNGVYASME----- 452
DB 399 YGIGELASADQVYFYFIDKTGAT-----LNSGCGN-VIASERPMMKRFYDVTY 450
QY 453 -----HCNDFME--LG--TEAISTLGRVGDFFMCTDPS--GDPNGTFW-----L 489
DB 451 YVWKEVHIDGFRDQGLDKKTMLEVERALKHIDPTIILYGEPMWG--WGAPIRFGKSUV 509
QY 490 QGCHWYHCANDSLMNGNFTHPDMFQSTHPCAAFHAAASRA--ISGPIYVSDS-----Y 542
DB 510 AGTH--VAAFNDE-----FRDAIRGSVFPNPSVKGFMVG--YGETKIKIRGVY 553
QY 543 GKHNFDLLKKLVLPDGSILRSEYVALPTRDCLFEDPLHNGETM-----LKITV-- 589
DB 554 GSIYNT-----DGKLKLS--FALDPRETTIYVACHONHILMDKNYLAARADKKKEVT 602
QY 590 -----NLNKFYGVIGAFNCG-----GC--WCRETRRNCFSQYSKRVTSKTPKDIEMH 637
DB 603 EELKNAQKLAGAI--LLTSGQVFLHGGDFCFTTNFND-----N 641
QY 638 SGENPISIEG 647
DB 642 SYNAPISING 651

RESULT 9
US-09-771-161A-173
Sequence 173, Application US/09771161A
Patent No. US20020110811A1
GENERAL INFORMATION:
APPLICANT: LEVINE, et al.
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
FILE REFERENCE: 802620-2005.1
CURRENT APPLICATION NUMBER: US/09/771,161A
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: 09/724,676
PRIOR FILING DATE: 2000-11-28
PRIOR APPLICATION NUMBER: 136776
PRIOR FILING DATE: 2000-06-15
PRIOR APPLICATION NUMBER: 135619
PRIOR FILING DATE: 2000-04-12
NUMBER OF SEQ ID NOS: 273
SOFTWARE: PatentIn version 3.0
SEQ ID NO: 173
LENGTH: 344
TYPE: PRT
ORGANISM: Homo sapiens
US-09-771-161A-173

Query Match 2.3%; Score 95.5; DB 10; Length 344;
Best Local Similarity 27.0%; Pred. No. 1.5;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDGPRVFLPIVEGPFRTSIOPGDDDFVDVCV--ESGSSKVVDASFRLM 173
DB 184 ESSVSILQ---GARSPFCMPICE-----ISRGTINFSEELKIGSGGCGCYRAAMRNT 234
QY 174 LVL-----HAGDDPFALVEAMKIVRTHLGTFRLLLEKTPPGIVDKFGWCMTDAFYLVH 228
DB 235 YVAVKRLKENADLEWYAVKQSEFLTEVEQLSRF-----HPNIVDPAGYCAONGFYCLVY 288
QY 229 ---PGVIEGVHRLVDGCGRP 246
DB 289 GLPLNGSLDRHLHCOTQACRP 309

RESULT 10
US-09-773-753-5
Sequence 5, Application US/09773753
Patent No. US20020099177A1
GENERAL INFORMATION:
APPLICANT: NI, JIAN
FENG, PING
MUZIO, MARTA
DIXIT, VISHVA M.
TITLE OF INVENTION: HUMAN IRAK-2
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVENUE, N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/773,753
FILING DATE: 02-Feb-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/980,060
FILING DATE: <unknown>

```

;
; ATTORNEY/AGENT INFORMATION:
; NAME: STEFFE, ERIC K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.1010000/EKS/AJK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 712 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: No. US20020099177A1 Relevant
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-773-753-5

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
DB 184 ESSVSLQ---GARPSFPCPLCE-----ISRGTHNFSEELKIGEGFGCVRAVWRNT 234
QY 174 LYL-----HAGDDPFAIVKRAMKIVRTHLGTFRLLLEKTPPGIIVDKFGWCTWDAFYLTVA 228
DB 235 VYAVKRLKENADLEMTAVKOSFLEVEQLSRFR-----HPNIVDPRAGYCAQNGFCYCLVY 288
QY 229 ---POGVIEGVRHLVDGCGCP 246
DB 289 GFLPNSGLEDRHLHCOTQACPP 309

RESULT 11
US-09-771-161A-262
; Sequence 262, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 262
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-262

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
DB 184 ESSVSLQ---GARPSFPCPLCE-----ISRGTHNFSEELKIGEGFGCVRAVWRNT 234
QY 174 LYL-----HAGDDPFAIVKRAMKIVRTHLGTFRLLLEKTPPGIIVDKFGWCTWDAFYLTVA 228
DB 235 VYAVKRLKENADLEMTAVKOSFLEVEQLSRFR-----HPNIVDPRAGYCAQNGFCYCLVY 288
QY 229 ---POGVIEGVRHLVDGCGCP 246
DB 289 GFLPNSGLEDRHLHCOTQACPP 309
```

```

;
; RESULT 12
US-09-771-161A-263
; Sequence 263, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 263
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-263

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
DB 184 ESSVSLQ---GARPSFPCPLCE-----ISRGTHNFSEELKIGEGFGCVRAVWRNT 234
QY 174 LYL-----HAGDDPFAIVKRAMKIVRTHLGTFRLLLEKTPPGIIVDKFGWCTWDAFYLTVA 228
DB 235 VYAVKRLKENADLEMTAVKOSFLEVEQLSRFR-----HPNIVDPRAGYCAQNGFCYCLVY 288
QY 229 ---POGVIEGVRHLVDGCGCP 246
DB 289 GFLPNSGLEDRHLHCOTQACPP 309

RESULT 13
US-09-771-161A-264
; Sequence 264, Application US/09771161A
; Patent No. US20020110811A1
; GENERAL INFORMATION:
; APPLICANT: LEVINE, et al.
; TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES
; FILE REFERENCE: 802620-2005.1
; CURRENT APPLICATION NUMBER: US/09/771,161A
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/724,676
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: 136776
; PRIOR FILING DATE: 2000-06-15
; PRIOR APPLICATION NUMBER: 135619
; PRIOR FILING DATE: 2000-04-12
; NUMBER OF SEQ ID NOS: 273
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 264
; LENGTH: 712
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-771-161A-264

Query Match
Best Local Similarity 27.0%; Score 95.5; DB 10; Length 712;
Matches 38; Conservative 16; Mismatches 62; Indels 25; Gaps 6;

QY 116 ETQIVILEKSDSGRPYVFLPIVEGPFRTSIQPGDDFVDVCV--ESGSSKVVDAFSRSM 173
```

```
Db 184 ESSVSLLQ---GARPSPFCWPLCE-----ISRGTHNFSSEELKIGGCGCYVAVAMNMT 234
Oy 174 LVL-----HAGDDPFLVKEAMKIVRTHLGFRLLEKTPGIVDKFGMCTWDAFLTVH 228
Db 235 VVAVVRKLEKNADLEWVAVQSFLEVEQLSRFR-----HPIVDPAGYCAQNGFICLYV 288
Oy 229 ---PGVIGEVRRHLVDGCGPP 246
Db 289 GFLPNGLSEDLRHLCQTOACPP 309

RESULT 14
US-10-101-464A-954
; Sequence 954, Application US/10101464A
; Publication No. US20030046728A1
; GENERAL INFORMATION:
; APPLICANT: Strabala, Timothy
; APPLICANT: Nieuwenhuizen, Nicolaas
; TITLE OF INVENTION: Compositions Isolated from Plant Cells
; FILE REFERENCE: 11000.1020C2
; CURRENT APPLICATION NUMBER: US/10/101,464A
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/704,302
; PRIOR FILING DATE: 2000-11-01
; PRIOR APPLICATION NUMBER: 09/228,986
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/162,866
; PRIOR FILING DATE: 1999-11-01
; PRIOR APPLICATION NUMBER: PCT/US00/00724
; NUMBER OF SEQ ID NOS: 989
; SOFTWARE: FastSeq for Windows version 4.0
; SEQ ID NO 954
; LENGTH: 1021
; TYPE: PRT
; ORGANISM: Eucaalyptus grandis
US-10-101-464A-954

Query Match 2.2%; Score 94; DB 9; Length 1021;
Best Local Similarity 20.7%; Pred. No. 11;
Matches 139; Conservative 97; Mismatches 217; Indels 220; Gaps 42;

Oy 164 KVVDAFSRMLYLAHAGDDPFLVKEAMKIVRTHLGFRLLEKTPGIVDKFGMCTWDAF 223
Db 152 RVLDDLNNNM---TGDLPVA--VAEMPOLRHLHLCG--NFGGIRP-----EYC--RMEFI 199
Oy 224 -VLTVH-----PO--GVIEGVRHL-----VDGCGPPGLVLIDGWSQISGHDSPI 265
Db 200 EYLAVSGDELTAIPPEIGNLFTLRILGYVNSYDGLRP-----EIGNLSQLY 249
Oy 266 TEGMNOYVAGGEMPCRLKFOENYKFRDYN-----FKATGPRAGOKGM-----KAFLD 315
Db 250 RLDAAACGSGE--IPEIKLLOMLDTLFLQVNGFAGSLPAEIGYLSLSDLSNNMFAG 308
Oy 316 ELKGEKTEVHYVVMALCGYMGRLPO--VPGDPEARVIO-----PYLSPLC-----OM 363
Db 309 EIPESLSQAKNLTLLHLFRNKLNGESPFIADLPLOVQLWMNNTGSPGLGKNGHL 368
Oy 364 TMEDLAVDKIVLHKVGLVPRPEKAEMYEGSLHAHLEKVGIDYKIDVILHLEMLCEDYGR 423
Db 369 QIVDLSSNKL---TGLTPD-----LC--YGNO 391
Oy 424 VDLAKAYYKAMTSSINKHKGNGVIAEMHCNFMFLGTEAISLGRVGDGDFWCTDPSDP- 482
Db 392 LQTLILSNVLLGPIRESL---GKCRSLEKI-----RMGGNYL-----NGSI 430
Oy 483 PNCFTWLOGCHVNCANDSLMGNFTIHPDMFQSTHPCAAFAHAASRAISGCEPIYSD-- 540
Db 431 PRGLFGLPELNOVE--FQDNLLVGEFPPVD-----DSIALKLGQITITSNNK 474
Oy 541 -----SVGKHNFLLKLLVLPDGSILNSEYVALPTRDCLFEDPLHNGETMLKI--WNLN 592
```

```
Db 475 LCGSLPPTIG--NFSGVQKLLD--DGNFSGO---IPPE-----IGRLQOLAKIDFSSN 521
Oy 593 KFTGVGARNCGGGGCRTRNOCFSQYSKRVSTNTKDIEMH--SEENISIEGYVT 650
Db 522 RYSGPIPA-----QISQC-----KLTTF---VDLSGNLSGEINETIGMRI 560
Oy 651 FALYTYOAKKLLISRPDDLDIALDPFEFLITVSPVTKLIQTSLHFAPIGLVNMLNTSG 710
Db 561 LN-----YLNLSNNLSGSIIP-----SISTMOSL--TSVDFSTNNLSGLVPGTG 603
Oy 711 AIOQVDYDDDLSSVEIGVKGCEMRVFAASKPRACRIDEDYGFKYDDQMVVVOYVPMPI 770
Db 604 QFSYFNVTSTFLNPPEL---CGPYL-----GPCK--DDEMTGIHQ-----AHVKGPL 644
Oy 771 DSSSGISVIEVL 783
Db 645 SASVKLLIVIGLL 657

RESULT 15
US-09-759-508B-2
; Sequence 2, Application US/09759508B
; Publication No. US20020182599A1
; GENERAL INFORMATION:
; APPLICANT: Fishman, Mark C.
; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease
; FILE REFERENCE: 00786/381002
; CURRENT APPLICATION NUMBER: US/09/759,508B
; CURRENT FILING DATE: 2001-01-12
; PRIOR APPLICATION NUMBER: US 60/175,787
; PRIOR FILING DATE: 2000-01-12
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 26926
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-759-508B-2

Query Match 2.2%; Score 93; DB 9; Length 26926;
Best Local Similarity 19.9%; Pred. No. 2,3e+03;
Matches 97; Conservative 61; Mismatches 136; Indels 194; Gaps 28;

Oy 7 NGSNNVVF--DGLND-----MSSPFAIDGSDFTVNGHSFLSDVPEINVAS 50
Db 13413 DGGSRIIGYVDFLTIEENKQRYMKLSLQYSAKDLTEGKETIFR---VSAENENGEGT 13468
Oy 51 PSPYTSIDKSPVSVGCFVGFDASEPPDSRHVVSIGLKLDIRFM---SIFRFKV----- 99
Db 13469 PSEITIVVARD-----DVVAPD---LDLKGLPDLCYLAKENSFRKLPKIGKPA 13514
Oy 100 ---WNTTHVGRNGD--LESETQIYI-----LEKSDSGRPYVFLPLPIVEGP 141
Db 13515 PVSVM-----KKGEDPLATDTRVSVESAVNTLLIYVDCQSDPAK--YTTTLKKNVAGT 13566
Oy 142 -----FRTSIOPGDDDFVDCVES-----GSSKYVDASFRLMYLAHAD 180
Db 13567 KEETSIKAVNGKRGITGTGKIKDEYVAELMALKMAPPKDDGSEITN-----YILEKR 13619
Oy 181 DPF-----ALVKEAMKIVRTHLG--TPRLLEKTPPGIVDKFGMCTWDAFLTVH 228
Db 13620 DSVNNKNWVCASAVQKTTFRVTRRLHEGMETPRVSAE-----N 13657
Oy 229 POCVITGV-----KH-----LVDGCGPPGLVLIDGWSQISGHDSPIETEGMNOYVAGEQ 278
Db 13658 KYGVGGLSEPIVARHPEDVPDAPPPNIV-----DVRHDSVSLTWTDPKKT--GCGSP 13709
Oy 279 MPCRLKFOEN-----YKFRDYNPKATGPRAGOKGMKAFLDELKGEFTVGH 326
Db 13710 ITGYHLHEFERNSLMLKRRANKPIRMKDF---KYVTLTEG-----LEVEFPVM-- 13754
Oy 327 VVVMHALCGYMGRLPOVGLPEARVIOPLSPGLQMTMEDLAVDKIVLHKVGLVPRPEKA 386
```

Db 13755 -----AINLAGVGKPSLPSEPVPV-ALDPIDPPGKP-----EYINITRNSVTLLI---WT 13798
 QY 387 EEMVEGLH 394
 Db 13799 EPKIDGGH 13806

Search completed: April 4, 2003, 11:18:14
 Job time : 47 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 14:59:10 ; Search time 7021 Seconds

(without alignments)
9749,295 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407

Perfect score: 2352

Sequence: 1 ATGGCTCTAGATTAA...CGTTATCGAGTACTGTTT 2352

Scoring table: IDENTITY NUC Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0% Maximum Match 100%

Listing first 45 summaries

Database :
1: GenBank:
2: gb-ba:
3: gb-in:
4: gb-om:
5: gb-ov:
6: gb-pat:
7: gb-ph:
8: gb-pl:
9: gb-pr:
10: gb-ro:
11: gb-sts:
12: gb-sy:
13: gb-un:
14: gb-vi:
15: em-ba:
16: em-fun:
17: em-hum:
18: em-in:
19: em-mu:
20: em-om:
21: em-or:
22: em-ov:
23: em-pat:
24: em-ph:
25: em-pl:
26: em-ro:
27: em-sts:
28: em-un:
29: em-vi:
30: em-htg-hum:
31: em-htg-inv:
32: em-htg-other:
33: em-htg-mus:
34: em-htg-pln:
35: em-htg-rod:
36: em-htg-mam:
37: em-htg-vrt:
38: em-sy:
39: em-htg-hum:
40: em-htg-mus:
41: em-htg-other:

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
2352	100.0	2517	6	E15707	E15707 Cucumis sat
2352	100.0	2569	6	E25430	E25430 Rafinose s
2352	100.0	2569	8	AF073744	AF073744 Cucumis s
1041.4	44.3	2690	6	E36417	E36417 Rafinose s
976.4	41.5	2498	6	AR182382	AR182382 Sequence
975.4	41.5	2690	6	E37133	E37133 Rafinose s
966.8	41.1	2343	6	BD010263	BD010263 Mutant pr
966	41.1	2497	6	E24424	E24424 Rafinose s
956.6	40.7	2671	6	PSA426475	PSA426475 Pisum sat
956.6	40.7	2746	6	E24423	E24423 Rafinose s
950.6	40.4	2395	8	AY081645	AY081645 Arabidops
950.6	40.4	2718	8	AY062781	AY062781 Arabidops
886.8	37.7	1762	6	E24425	E24425 Rafinose s
866.4	36.8	1762	6	E36418	E36418 Rafinose s
406.2	17.3	2727	8	PSA311087	PSA311087 Pisum sat
400.4	17.0	2592	8	SS1344091	SS1344091 Stachys s
387.6	16.5	2889	8	AME487030	AME487030 Alonsoa m
387	16.5	3073	8	VANI9024	VANI9024 Vigna angl
364	15.5	135295	8	AP003282	AP003282 Oryza sat
364	15.5	175174	8	AP003339	AP003339 Oryza sat
303.8	12.9	928	6	E30796	E30796 Rafinose s
301.6	12.8	2714	6	AY090237	AY090237 Arabidops
275.8	11.7	2780	6	E25448	E25448 Rafinose s
274.2	11.7	2738	8	AY050772	AY050772 Arabidops
263.8	11.2	86263	8	AB006702	AB006702 Arabidops
252.2	10.7	2669	8	BOBUTIN	BOBUTIN B. oleracea
237.6	10.1	2673	8	PM133148	PM133148 Petrea am
228	17.0	4	8	AC007138	AC007138 Arabidops
228	17.0	4	8	ATCHRIV5	ATCHRIV5 Arabidops
169.6	7.2	993	6	E24426	E24426 Rafinose s
157.8	6.7	101458	8	F5024	F5024 Arabidops
130.8	5.6	1487	8	CAR271668	CAR271668 Cicer ari
111.8	4.8	120868	2	AP003748	AP003748 Oryza sat
97	4.1	122095	8	AP002854	AP002854 Oryza sat
90.8	3.9	253	6	E25451	E25451 Rafinose s
81.4	3.5	87503	8	ATT8H10	ATT8H10 Arabidops
78.4	3.3	850	8	AF159378	AF159378 Arabidops
71.8	3.1	4788	8	BLXSTP1A	BLXSTP1A Hordeum vul
70.6	3.0	125020	9	AF429315	AF429315 Homo sapi
66	2.8	90767	2	AC002328	AC002328 Genomic s
64.4	2.7	129510	8	H0510A06	AL442104 Oryza sat
64	2.7	154909	2	OSJN00047	AL606514 Oryza sat
64	2.7	162317	2	AP005174	AP005174 Oryza sat
56	2.4	1071	8	CASIMBIBP	AP005148 Oryza sat
					X95875 C. artefactum

ALIGNMENTS

RESULT 1
LOCUS E15707 2517 bp DNA linear PAT 28-JUL-1999
DEFINITION Cucumis sativus mRNA for raffinose synthase, complete cds.
ACCESSION E15707
VERSION E15707.1 GI:5710390
KEYWORDS JP 1998084973-A/1.
SOURCE JP 1998084973-A/1.
ORGANISM Cucumis sativus
unclassified.
REFERENCE 1 (bases 1 to 2517)
AUTHORS Oosumi, C., Nozaki, J. and Kida, T.
TITLE RAFFINOSE SYNTHASE GENE, PRODUCTION OF RAFFINOSE AND TRANSFORMED
JOURNAL PATENT: JP 1998084973-A 1 07-APR-1998;
AJINOMOTO CO INC

COMMENT OS Cucumis sativus (cucumber)
PN JP 1998084973-A/1
PD 07-APR-1998
PE 28-APR-1997 JP 1997111124
PR 26-APR-1996 JP 96P 107682, 26-JUL-1996 JP 96P 198079 PI
OOSUMI CHIEKO, NOZAKI JINJI, KIDA TAKAO
PC C12N15/09,A01H5/00,C12M9/00;
CC strandedness: Double;
CC topology: Linear;
FH Key Location/Qualifiers
FH source 1..2517
FT /organism='Cucumis sativus'
FT /tissue_type='leaf'
FT CDS 56..2410
FT /product='raffinose synthase'.
FEATURES
source Location/Qualifiers
1..2517
BASE COUNT 613 a 524 c 642 g 738 t
ORIGIN
Query Match 100.0%; Score 2352; DB 6; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 ATGGCTCTAGTTTAAATGGTGGCTCAACGCTAGTTTCATTGATGGCTTAATGAC 60
Db ATGGCTCTAGTTTAAATGGTGGCTCAACGCTAGTTTCATTGATGGCTTAATGAC 115
QY ATGTCGTCAACCGTTTGCATTCGACGAGTGGATTCTACTGTGACGCTATTCGTTTCTG 120
Db ATGTCGTCAACCGTTTGCATTCGACGAGTGGATTCTACTGTGACGCTATTCGTTTCTG 175
QY 121 TCGGATGTCCTGAGAACATGTTGCTTCCTTCCTCCGACACTTCGATGAGCAAGTCC 180
Db 176 TCGGATGTCCTGAGAACATGTTGCTTCCTTCCTCCGACACTTCGATGAGCAAGTCC 235
QY 181 CCGGTTTCGGTTGGTTGCTTTGTTGGATTGACGCGTCGGAACCTGATAGCCGACATGTT 240
Db 236 CCGGTTTCGGTTGGTTGCTTTGTTGGATTGACGCGTCGGAACCTGATAGCCGACATGTT 295
QY 241 GTTTGCATTGGGAAGCTGAAGGATATTCGTTTATGAGTATTTTCAGGTTTAAAGTTTGG 300
Db 296 GTTTGCATTGGGAAGCTGAAGGATATTCGTTTATGAGTATTTTCAGGTTTAAAGTTTGG 355
QY 301 TGGACTACACACTGGGTTGGTGGAAATGGTGGGGGATCTGAATCGGAGACTGATTTGG 360
Db 356 TGGACTACACACTGGGTTGGTGGAAATGGTGGGGGATCTGAATCGGAGACTGATTTGG 415
QY 361 ATCCCTTGAGAAGTCAAGTCTGGTGCACGCTATGTTTCTCTTCCTCCGATCGTTGAGGGA 420
Db 416 ATCCCTTGAGAAGTCAAGTCTGGTGCACGCTATGTTTCTCTTCCTCCGATCGTTGAGGGA 475
QY 421 CCGTTTCCGAACCTCCGATTAGCCTGGGGATGATGACTTTTGTGATTTTGTGTCGAGAT 480
Db 476 CCGTTTCCGAACCTCCGATTAGCCTGGGGATGATGACTTTTGTGATTTTGTGTCGAGAT 535
QY 481 GGTTCGTCGAAGTGTGATGATGATGATCGTCCGAAGTATGTGATTTGATGCTGATGAT 540
Db 536 GGTTCGTCGAAGTGTGATGATGATGATCGTCCGAAGTATGTGATTTGATGCTGATGAT 595
QY 541 GATCCGTTTGACATTGTTAAAGAGGCGATGAAGATCGTAGAGACCCATCTTGAACCTTTT 600
Db 596 GATCCGTTTGACATTGTTAAAGAGGCGATGAAGATCGTAGAGACCCATCTTGAACCTTTT 655
QY 601 CGCTTGTGGAGAGAAAGCTCCACAGAGTATCGTGAACAAATTCGTTGGTGTGCACGTGG 660
Db 656 CGCTTGTGGAGAGAAAGCTCCACAGAGTATCGTGAACAAATTCGTTGGTGTGCACGTGG 715
QY 661 GACGGCTTTTACCTAAGGTTTCATCCACAGGGCGTAAATGAAGGGGTGAGGATCTCGTC 720
|||||

Db 716 GACGGCTTTTACCTAAGGTTTCATCCACAGAGGGCGTAAATGAAGGGCTGAGGCACTCTGTC 775
QY 721 GACGGCGTTTGTCTCCCGGTTTGTAGTCTTAATGACAGATGTTGGCAATTCATCGGACAC 780
Db 776 GACGGCGTTTGTCTCCCGGTTTGTAGTCTTAATGACAGATGTTGGCAATTCATCGGACAC 835
QY 781 GATTGGATCCCATCAACCAAGAAGATGAACCAACCGTCCGCGGAGCAAAATGCCC 840
Db 836 GATTGGATCCCATCAACCAAGAAGATGAACCAACCGTCCGCGGAGCAAAATGCCC 895
QY 841 TGGCGTCTTTTGAATTTCCAAAGAAATTACAAATTCCTGACACTACTGCAATCCCAAGGCC 900
Db 896 TGGCGTCTTTTGAATTTCCAAAGAAATTACAAATTCCTGACACTACTGCAATCCCAAGGCC 955
QY 901 ACCGGCCCCCGACCGGCGCAGAGGGGATGAAGCGCTTTATAGATCACTCAAGAGAG 960
Db 956 ACCGGCCCCCGACCGGCGCAGAGGGGATGAAGCGCTTTATAGATCACTCAAGAGAG 1015
QY 961 TTTTAACTGTGAGAGCATGTTATGTTTGGCATGCTTTGTGATATTTGGGTGGCCTT 1020
Db 1016 TTTTAACTGTGAGAGCATGTTATGTTTGGCATGCTTTGTGATATTTGGGTGGCCTT 1075
QY 1021 CCGCCCGAGGTGCTGCTGCTTGGTGGAGCAGCTGTGATTCAGCACTGCTTTCACAGGG 1080
Db 1076 CCGCCCGAGGTGCTGCTGCTTGGTGGAGCAGCTGTGATTCAGCACTGCTTTCACAGGG 1135
QY 1081 CTGCAGATGACGATGAGGATTTGGCGGTGATGAATGTTCTTCAATGAAGTCGGGCTG 1140
Db 1136 CTGCAGATGACGATGAGGATTTGGCGGTGATGAATGTTCTTCAATGAAGTCGGGCTG 1195
QY 1141 GTCCCGCGGAGAGGCTGAGAGATGTACGAAGAACTTCATGCTCATTTTGGAAAAAGTT 1200
Db 1196 GTCCCGCGGAGAGGCTGAGAGATGTACGAAGAACTTCATGCTCATTTTGGAAAAAGTT 1255
QY 1201 GGGATGACGGGTGTTAAGATTTGACGTTATCCACCTTFTGGAAGTGTGTTGTAAGCTAT 1260
Db 1256 GGGATGACGGGTGTTAAGATTTGACGTTATCCACCTTFTGGAAGTGTGTTGTAAGCTAT 1315
QY 1261 GGAAGGAGAGTGATTTGGCAAAAGGCAATTTACAAAGCAATGCCAAATCAATAATATAA 1320
Db 1316 GGAAGGAGAGTGATTTGGCAAAAGGCAATTTACAAAGCAATGCCAAATCAATAATATAA 1375
QY 1321 CATTTTAAAGGAATGAGTCAATGCAAGTATGGAACATTTGTAACACTTTCATGTTCTT 1380
Db 1376 CATTTTAAAGGAATGAGTCAATGCAAGTATGGAACATTTGTAACACTTTCATGTTCTT 1435
QY 1381 GGCACGGAAGCTATCTCTGTCGTCGTTGGTGGATGATGATTTGGTGGACGACCCTT 1440
Db 1436 GGCACGGAAGCTATCTCTGTCGTCGTTGGTGGATGATGATTTGGTGGACGACCCTT 1495
QY 1441 GGTGATCCAAAGCGTACGTTTGGCTCAAGAGATGTCACATGGTTTCATTTGTCACAGAC 1500
Db 1496 GGTGATCCAAAGCGTACGTTTGGCTCAAGAGATGTCACATGGTTTCATTTGTCACAGAC 1555
QY 1501 AGCTTGTGATGGGGAACCTTCAACCTCGACTGGGATATGTTCCAAATCCACCCACT 1560
Db 1556 AGCTTGTGATGGGGAACCTTCAACCTCGACTGGGATATGTTCCAAATCCACCCACT 1615
QY 1561 TGTGCGCCTTCCATCTCTCTCTGAGGCATCTCGTGGGCGCCGATCTATGTTAGTGA 1620
Db 1616 TGTGCGCCTTCCATCTCTCTCTGAGGCATCTCGTGGGCGCCGATCTATGTTAGTGA 1675
QY 1621 TCTGTGGGAAGCATTTGATCTTCTGAAAAAACTAGTCTTCCTGATGATGATGATC 1680
Db 1676 TCTGTGGGAAGCATTTGATCTTCTGAAAAAACTAGTCTTCCTGATGATGATGATC 1735
QY 1681 CTTCGAAGTACTATGATCACTCCGAGTCCGAGATTTGTTTGTGGAACACCTTTGGAT 1740
Db 1736 CTTCGAAGTACTATGATCACTCCGAGTCCGAGATTTGTTTGTGGAACACCTTTGGAT 1795
QY 1741 AATGGAAGAACTATGCTTAAATTTGGAATCTCAACAAATTCAGTGAAGTATTTGATGCA 1800
Db 1796 AATGGAAGAACTATGCTTAAATTTGGAATCTCAACAAATTCAGTGAAGTATTTGATGCA 1855
|||||

Accession	Source	Organism	Reference Authors Title	Journal	Comment
QY 1801	TTCAACTGCGCAAGAGAGAGATGTGTGTGTGAGACAGCCCGCAACCAATGCTTTTTCACAA	1860			
Db 1856	TTCAACTGCGCAAGAGAGAGATGTGTGTGTGAGACAGCCCGCAACCAATGCTTTTTCACAA	1915			
QY 1861	TACTTCAAAAGAGTACATCCAAAACCTAACCCAAAAGACATAGATGGCAGCAGTGGAGAA	1920			
Db 1916	TACTTCAAAAGAGTACATCCAAAACCTAACCCAAAAGACATAGATGGCAGCAGTGGAGAA	1975			
QY 1921	AACCCATCTCTGATTTGAAAGGCGTTTAAACCTTTGGCGTTTACCTCTATTCAGCCAAAAA	1980			
Db 1976	AACCCATCTCTATTTGAAAGGCGTTTAAACCTTTGGCGTTTACCTCTATTCAGCCAAAAA	2035			
QY 1991	CTTATCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTTCGAATTGCG	2040			
Db 2036	CTTATCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTTCGAATTGCG	2095			
QY 2041	CTCATCACTGTTTTCACAGTACCCAACTCATCCAAACTTCTCTACACTTTGGCCCAATT	2100			
Db 2096	CTCATCACTGTTTTCACAGTACCCAACTCATCCAAACTTCTCTACACTTTGGCCCAATT	2155			
QY 2101	GGGCGGTGAACATGCTTAAACCTAGTGGAGCCATCCAACTGTGGACTATGACATGAC	2160			
Db 2156	GGGCGGTGAACATGCTTAAACCTAGTGGAGCCATCCAACTGTGGACTATGACATGAC	2215			
QY 2161	CTAAGCTCAGTCGACATTTGGTGTCAAAGGGTGTGTGAGATGCGAGTATTTGCATCGAAA	2220			
Db 2216	CTAAGCTCAGTCGACATTTGGTGTCAAAGGGTGTGTGAGATGCGAGTATTTGCATCGAAA	2275			
QY 2221	AAACCAAGGGCTGTGCGATTTGATGTGGGAGATGTGGTTCAAGTATGATGACAGCA	2280			
Db 2276	AAACCAAGGGCTGTGCGATTTGATGTGGGAGATGTGGTTCAAGTATGATGACAGCA	2335			
QY 2281	ATGCGTGTGTTTCAAGTGCATGGCCAAATTGATCTTCATCGGGTGGCATTTTCGGTTATC	2340			
Db 2336	ATGCGTGTGTTTCAAGTGCATGGCCAAATTGATCTTCATCGGGTGGCATTTTCGGTTATC	2395			
QY 2341	GAGTACTTGTGT 2352				
Db 2396	GAGTACTTGTGT 2407				
RESULT 2					
LOCUS	E25430	2569 bp	DNA	linear	PAT 18-JUN-2001
DEFINITION	Raffinose synthase gene, process for producing raffinose and				
ACCESSION	E25430				
VERSION	E25430.1	GI:13024726			
KEYWORDS	JP 1999123080-A/1.				
SOURCE	Cucumis sativus.				
ORGANISM	Cucumis sativus.				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
AUTHORS	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots;				
TITLE	Rosidae; eurosids I; Cucurbitales; Cucurbitaceae; Cucumis.				
JOURNAL	Chieko, O., Jinji, N. and Takao, K.				
COMMENT	Raffinose synthase gene, process for producing raffinose and				
	transformed plant				
	Patent: JP 1999123080-A 1 11-MAY-1999;				
	AJINOMOTO CO INC				
	OS Cucumis sativus (Cucumber)				
	PN JP 1999123080-A/1				
	PD 11-MAY-1999				
	PE 24-OCT-1997 JP 1997292969				
	PR				
	PI CHIEKO OZUMI, JINJI NOZAKI, TAKAO KIDA				
	PC C12N15/09, A01H5/00, C12N9/00, C12R1.19, C12N15/00 CC				
	Strandedness: Double;				
	CC Topology: Linear;				
	EH Key				
	FT CDS				
FEATURES	Location/Qualifiers				
	56..2407.				

[illegible]

QY 961 TTTAAGACTGTGGAGCATGTTTATGTTTGGCATGCTTTGTGTGATATTGGGGTGGCTT 1020
 |||||
 Db 1016 TTTAAGACTGTGGAGCATGTTTATGTTTGGCATGCTTTGTGTGATATTGGGGTGGCTT 1075
 QY 1021 CGCCCGAGGTGCTTGGCTTGGCTGAGGACGTGTGATTCAGCCAGTGCCTTCCACAGG 1080
 |||||
 Db 1076 CGCCCGAGGTGCTTGGCTTGGCTGAGGACGTGTGATTCAGCCAGTGCCTTCCACAGG 1135
 QY 1081 CTGAGATGAGTGGAGGATTTGGCGGTGATTAAGATTGTTCTTCAATAGTGTGGCTG 1140
 |||||
 Db 1136 CTGAGATGAGTGGAGGATTTGGCGGTGATTAAGATTGTTCTTCAATAGTGTGGCTG 1195
 QY 1141 GTCCCGCGGAGAGGCTGAGAGATGTACGAAGGACTTATGCTCATTTGGAAAAAGTT 1200
 |||||
 Db 1196 GTCCCGCGGAGAGGCTGAGAGATGTACGAAGGACTTATGCTCATTTGGAAAAAGTT 1255
 QY 1201 GGGATCGACGCTGTTAAGATTGACGTTATCCACCTATTGGAGATGTTGTGTGAAGACTAT 1260
 |||||
 Db 1256 GGGATCGACGCTGTTAAGATTGACGTTATCCACCTATTGGAGATGTTGTGTGAAGACTAT 1315
 QY 1261 GGAAGGAGATGATTTGGCAAGGCAATTACAAAGCAATGACCAATCAATTAATAAA 1320
 |||||
 Db 1316 GGAAGGAGATGATTTGGCAAGGCAATTACAAAGCAATGACCAATCAATTAATAAA 1375
 QY 1321 CATTTTAAAGAAATGAGATCATTTGCAAGTATGGAACATTTGTAACGACTTCATGTTCTT 1380
 |||||
 Db 1376 CATTTTAAAGAAATGAGATCATTTGCAAGTATGGAACATTTGTAACGACTTCATGTTCTT 1435
 QY 1381 GGCACGGAAGCTATCTCTCTGTTGCTGCTGTTGGTATGATGATCTTTGGTGACAGACCCCTCT 1440
 |||||
 Db 1436 GGCACGGAAGCTATCTCTCTGTTGCTGCTGTTGGTATGATGATCTTTGGTGACAGACCCCTCT 1495
 QY 1441 GGTGATCCAAACGATGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
 |||||
 Db 1496 GGTGATCCAAACGATGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1555
 QY 1501 AGCTTGTGATGGGGAACCTTCATCCACCCTGACTGGGATGTTGTTCCAAATCCACCACCT 1560
 |||||
 Db 1556 AGCTTGTGATGGGGAACCTTCATCCACCCTGACTGGGATGTTGTTCCAAATCCACCACCT 1615
 QY 1561 TGTGCGCCCTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
 |||||
 Db 1616 TGTGCGCCCTTCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1675
 QY 1621 TGTGTTGGGAAGCATTAATTGATCTTCTGCAAAAACTAGTCTTCTGATGATGATC 1680
 |||||
 Db 1676 TGTGTTGGGAAGCATTAATTGATCTTCTGCAAAAACTAGTCTTCTGATGATGATC 1735
 QY 1681 CTTCGGAAGTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 |||||
 Db 1736 CTTCGGAAGTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795
 QY 1741 AATGGAAGAACTATGCTTAAGATTGGAATCTCAACAAGTTCATGGAGATTTGGTGA 1800
 |||||
 Db 1796 AATGGAAGAACTATGCTTAAGATTGGAATCTCAACAAGTTCATGGAGATTTGGTGA 1855
 QY 1801 TTCACTGCGCAAGAGAGAGATGCTGCTGAGACAGCGCCGAACCAATGCTTTTCACAA 1860
 |||||
 Db 1856 TTCACTGCGCAAGAGAGAGATGCTGCTGAGACAGCGCCGAACCAATGCTTTTCACAA 1915
 QY 1861 TACTCAAAAAGTACTGATCCTCAAAACTAACCCAAAAGACATAGAAATGGCAGATGGAGAA 1920
 |||||
 Db 1916 TACTCAAAAAGTACTGATCCTCAAAACTAACCCAAAAGACATAGAAATGGCAGATGGAGAA 1975
 QY 1921 AACCCATGCTCTTGAAGGCGTTAAACCTTTGGCTTACTCTATCAAGCCAAAAA 1980
 |||||
 Db 1976 AACCCATGCTCTTGAAGGCGTTAAACCTTTGGCTTACTCTATCAAGCCAAAAA 2035
 QY 1981 CTATATCTCTCCAAAGCCCTCTCAAGATCTTGAATAGCTCTTGAACCATTCGAATTCGAG 2040
 |||||
 Db 2036 CTATATCTCTCCAAAGCCCTCTCAAGATCTTGAATAGCTCTTGAACCATTCGAATTCGAG 2095

QY 2041 CTATACACTGTTTACCAGGAGTACCAACATCATCAAACTTCTACACTTTGCCCAATT 2100
 |||||
 Db 2096 CTATACACTGTTTACCAGGAGTACCAACATCATCAAACTTCTACACTTTGCCCAATT 2155
 QY 2101 GGGCTGTTGAACATGCTTAACTAGTGAAGCATTCATCTGTGACATATGACATGAC 2160
 |||||
 Db 2156 GGGCTGTTGAACATGCTTAACTAGTGAAGCATTCATCTGTGACATATGACATGAC 2215
 QY 2161 CTAACTCATGAGATTTGGTGTCAAAAGGTTGGTGTGAGATGAGATTTGATCGAAA 2220
 |||||
 Db 2216 CTAACTCATGAGATTTGGTGTCAAAAGGTTGGTGTGAGATGAGATTTGATCGAAA 2275
 QY 2221 AAACCAAGGCTTGTCTATTTGATGGGAGAGATTTGGTGTCAAGTATGATCAAGACAA 2280
 |||||
 Db 2276 AAACCAAGGCTTGTCTATTTGATGGGAGAGATTTGGTGTCAAGTATGATCAAGACAA 2335
 QY 2281 ATGGTGTGTTCAAGTGGCATGGCCATTTGATTTTCATGGGTGATTTGGTTATC 2340
 |||||
 Db 2336 ATGGTGTGTTCAAGTGGCATGGCCATTTGATTTTCATGGGTGATTTGGTTATC 2395
 QY 2341 GAGTACTGTTT 2352
 |||||
 Db 2396 GAGTACTGTTT 2407

RESULT 3
 AF073744 2569 bp mRNA 1linear PLN 05-JAN-1999
 LOCUS Cucumis sativus raffinose synthase (Rts) mRNA, complete cds.
 DEFINITION AF073744
 ACCESSION AF073744
 VERSION AF073744.1 GI:4106394
 KEYWORDS
 SOURCE Cucumis sativus.
 ORGANISM Cucumis sativus.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots;
 Rosidae; eustosids I; Cucurbitales; Cucurbitaceae; Cucumis.
 REFERENCE
 AUTHORS Ohsumi, C., Nozaki, J.
 TITLES 1 (bases 1 to 2569)
 JOURNAL Direct Submission
 Submitted (22-JUN-1998) Central Research Laboratories, Asijiomoto
 Co., Inc., 1-1, Suzuki-cho, Kawasaki-ku, Kawasaki 210-8681, Japan
 FEATURES
 source
 1..2569
 location/Qualifiers
 /organism="Cucumis sativus"
 /db_xref="taxon:3659"
 /issue_type="leaf"
 /dev_stage="mature"
 1..2569
 /gene="Rts"
 56..2410
 /gene="Rts"
 /EC_number="2.4.1.82"
 /codon_start=1
 /product="raffinose synthase"
 /protein_id="AAD02832.1"
 /translation="MAPSEKNGSNVVSFDGLNDMSFPAIDSGFTVNGHSLSLSDVP
 ENIVASPSPTYSIDKSPVSGCFVGEDESEPSRHVSIGLKLDIREMSIPEFKVMT
 THWVRNGDLESEFOIYLKESDSDGPRVFLPIVEGPFRTSIOGDDDPVDCVES
 GSSKVVYDASFRLMLYLHAGDDPFALYKEMAKIVRIHLGTFRLLEKTPPGIYDKGWC
 TDWAVLYVHPOGVLEGRVHLVDGSCPRLVILIDGWSISIGHSDPITREGNQTVAG
 EQMPRLKLFQENYKFRDYVNPVKATGPRAQSGKMAFIDELKGEFKTEVHYVMIALK
 GYWGRLRQVPLRPARVLOPVLSGLQMTMEDLAVDKIVLHKVGLVPEKAEEMVEG
 LHAHLEKVGIDGVKIDVHILHLEMLCEDYGGRRDLAKAYKAMTKSINRHPKNGVYIAS
 MEHCNDPMLTGLEALSLRGVDEPCTPSGDPNCTPMIOGCHMYHCANDSIAMNENFI
 HPDWMFOSTHPCAFAHAASRAISGCPITYVSDSVGKHNFIDLKTLVLPDGLIRSEYV
 ALPTDCLFEEDPLHNGETMLKIMNLNFKTGVYGAFCNCGGCRBTRRNOCFQYSKR
 VTSKTPKDIEMHSGENPISLEGVTFALYLYQAKLILSKPSQDLIDALDPFELI
 TVSPVTKLIQTSILNHPAGIYVLMNTSGAISVDVDDLSVSEIGVKSGCEMRVASK
 KPRACRIGEDVGEFKYDDQDMVYVNPPIIDSSSGISIVIEYLP"
 BASE COUNT 641 a 527 c 643 g 755 t 3 others
 ORIGIN

Oy	1388	AAGATCTCCTTGGTCGTGTTGGTAGTGACTTTGGTGGACGAGAACCCCTCGTGTATC	1447
Db	1518	AAGCATCTCTCTAGAGTCGTCTCGGTGTATGACTTTTGGTGCACGGATTCATCTAGCGGCA	1577
Oy	1448	CMAACGCTAGCTTTTTGGCTCCAGAAGATTGCATGATGGTTCATTTGGTCCAAAGCACAGCTGT	1507
Db	1578	TAAAGGGAACATATTGGCTGGACAAGATGCCAATGGTCTACTGTGTCCTTCAACAAGATCTTT	1637
Oy	1508	GGATGGGGAACCTTCATCCACCCTGACTGGGATPATGTTCAAATCCACCCACCTTGTGGCG	1567
Db	1638	GGATGGGAAAATTTTCATCCAGAGCTGATTGGAGCATGTTCATGCTCCACACATCCTTGTGGCTG	1697
Oy	1568	CCCTTCATGCTGGCCCTCTGACCATCTCTGGTGGCCGATCTATGTTAGTATTCGTGGG	1627
Db	1698	AGTTTCATGCTGGCTTCTGTCTCCATCTCTCGGTGGGCCCATTTTACATCCAGCATGTGTGGG	1757
Oy	1628	GAAAGCATTAACCTTTGATCTTCTGAAAAAATAAGTCCCTCGATGAGATGCATCTTGGAA	1687
Db	1758	GCCAGCACAGATTTCGATCTCTTTGAAGGACATCTGTCTTCCCTGACGGTTCCGATTTTGAAGT	1817
Oy	1688	GTGACTACTATGCACTCCGCACTCGCGATGTGTGTGTTGAAGACCCCTTTCGATATAGAG	1747
Db	1818	GTGAGCACTATGCACTCCCAACTCGTGCACCGCTCTCTTTGAAGACCCCTTCATGATGGCA	1877
Oy	1748	AAACATGCTTAAAGTTTGGATCTCACCAATTCACCAATTCAGTGGATGATTTGGTTCATTA	1807
Db	1878	AAACATGCTCTCAAGATTTTGGAACTTGAACCAAGTACATGGAATTTTGTGACATTTCAACT	1937
Oy	1808	GCCAAGGAGGAGATGATGTGTGTGAGACACGCCGCAACCAATGCTTTTCACAATACTCAA	1867
Db	1938	GCCAAAGGAGGAGATGTGTGACAGAAACCAGCAACCAATGCTTTCTCCCATATGCCTTA	1997
Oy	1868	AACGATGACATCCAAAATCTAACCCAAAAGACATGAATGSCACAGTGGAGAAAACCTTA	1927
Db	1998	ACACGTAAACCGCCACACAATAATCCTAAGACAGTTGAATGGAACAGTGGAAACACCCAA	2057
Oy	1928	TCTCTATTGAAGGCGTTAAACCTTTGGCGCTTACCTCTATCAAGCCAAAACCTTAACC	1987
Db	2058	TCTCGGTGAAACCTGTAAAGATTTGCTTGTGTTCTTCTCAAGTCTTAAGAAAGCTTGGT	2117
Oy	1988	TCTCCAAACCCCTCTCAAMATCTTGACATAGGCTCTTGACCCCATTCGAATTCAGAGCTATCA	2047
Db	2118	TGTCTGGAACCAACGATATATCTCGAGATCACTTTTGGACCTTTCAAGTTTGAAGCTAATCA	2177
Oy	2048	CTGTTTCAACCAAGTGCACCAACTCATCCAACCTCTCTACACTTGGCCCCAATTTGGCGCT	2107
Db	2178	CTGTCTCACCAAGTTGTCACTATTGAGGGTAGTTCAGTTTCTCCATTCGGATTTGG	2237
Oy	2108	TGACATATCTTAAACCTAGTSGAGCCATCCAAATCTGTGTGACTATGACGATAGCACTAAGCT	2167
Db	2238	TTTAACATCTTAAACCTAGCGGTGCAAAATTCGATCTTGTGTATCATATCAGGA-----AT	2291
Oy	2168	CAGTGCAGATTTGGTCTYCAAAAGGTTGGTGGATGATGCGAGTATTTTGCATCGAAAANACCA	2227
Db	2232	CCGTTGAGATTGGAATGTCTGTGTCTGGAGAGATTGAGGTTTATATGCATCAAGAAACCTG	2351
Oy	2228	GGGCTTGTCTGATTTGATGGGAGATGTTGGGTTCACGATATGATCAGACCAAAATGTTGG	2287
Db	2352	CGAGCTGCAAAATATGATGSTGAAGTTGTGAATTTGGATAG--AAAGATCAATAGTGA	2408
Oy	2288	TGGTTCAAGTGCATGCC	2306
Db	2409	TGGTTCAAGTGCCTTGGTCT	2427

RESULT 5	AR182382	AR182382	2498 bp	DNA	linear	PAT 20-APR-2002
LOCUS	AR182382	Sequence	23	from patent US 6337430.		
DEFINITION	AR182382	Accession				
VERSION	AR182382.1	GI:20225298				
KEYWORDS						

SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	FEATURES	BASE COUNT	ORIGIN	Query Match	Best Local Similarity	Matches 1507:	Conservative	Score 976.4:	DB 6:	length 2498:	Indels	Gaps
	Unknown.	Unclassified.	1 (bases 1 to 2498)	Ishige,F., Watanabe,E. and Oeda,K.	Plant promoters	Patent: US 6537430-A 23 08-JAN-2002;											
	source			Location/Qualifiers		1..2498											
						/organism="unknown"											
							651 a	521 c	673 g	653 t							

OY	57	TCACATGTCGACCGTTGGCAATGACGAGCGATGGATTTTCACTGTGACAGGCGCATTCGTT	116
Db	118	TAATTTGCCCTTGTGTCATTAACCCCTAGAAAGATCAATTTTCTCGCCMACGGCCACCCCTTT	177
OY	117	TCCTGTCGATGTTCCCTGAGAACATTGTTGCTTCTCTTCCTCCGTAC-----AC	164
Db	178	TCTCAGGGAAGTTCCCGGAAACATATAGTCAACCCCTTCAACCATTCGACGCCCAAGAGTAG	237
OY	165	TTTCGATAGACAAAGTCCCGGTTTCGGTTGGTTGTGCTTTTGTGGATTCGACGGCTCGAACC	224
Db	238	TAAAGAACAAACGAGGACGACGACGTCGTAGGTTGGTTCGTGGGCTTCACGCGGACGACGCC	297
OY	225	TGATAGCCGACATGTTGTTTTCATTGGATTGGGAAGCTGAAGAGATATTCGGTTATAGTATTTT	284
Db	298	CAGAAAGCCGACACGTGGCTTCCCTGGGGAAGCTCAGAGGAATATAAATTTATGACGTATTT	357
OY	285	CAGGTTTAAAGCTTTGGTGAGACTACACACTGGGTGGTGCAGAAATGGTGGGATCTTTCATTC	344
Db	358	CCGGTTTAAAGGTGTGTGGAGCCTACACTCGGTGGGTATACAACGGACACGAAGCTGGAGCA	417
OY	345	GGAGACTCAGATGTGTGATTCCTTGAGAAAGTCAG---ATTCTGGTGCAGCGTATGTTTCCCT	401
Db	418	CGAGACACAGATGATGCTTCTCGACAAAACACACCGCTCGAGACGCCCTTTGTGTTGATAT	477
OY	402	TCTTTCCGATCGTTGAGGGACCGTTTCCGAACCTTCGATTCAGCCTGGGGATGATGACTTTGT	461
Db	478	TCTCCGATTCCTCCAAAGCCTCGTTCCGAGCCCTCCGCAACCGGTTTGGATGATTAGCT	537
OY	462	CGATGTTGTGTGCGAGAGTGTTCCTCGAAAGTTGTTGATGCATCGTTCGGAATGATGTT	521
Db	538	GGACCTTTTCATGAGGAGGCGGTGCACACAGTCTCTGTGGCTCAGCTTTCGGGAGCGCTTT	597
OY	522	GTAATCTCATGCTGGTGATGATCCGTTTGCACCTGTTTAAAGGGGATGAAGATCGGTGAG	581
Db	598	ATTAGCTCCACGTGGCCCATGACCCCGATACATTGTTGCTTTAGAGAAAGCAATTAAGTCCCTTAG	657
OY	582	GACCCATCTTGGAACTTTTCCGTTTGTGGAGGAGAAAGACTCCACAGCAGATGCTGTGACAA	641
Db	658	GATGCATTTGGGGAGCGTTCAAGCTTCTCGAGGAGAAACCGGGCCAGTGATCATAGACAA	717
OY	642	ATTGCGTTGGTGCACGCTGGAGCGCGTTTTTACCTAACGTTTCATTCACAGCGCCTAATYAGA	701
Db	718	GTTTGGTTGGTGTACATGAGGAGCGCCTTTTACTTTGAAGGTGCATCCCTCCGTCGTGTGGGA	777
OY	702	AGGCGTGAAGGCACTTCGTCGACGAGGGGGTGTCTCCCGGTTTAGTCCCTAATGACAGCATGG	761
Db	778	AGGGGTGAAGGTTTGGTGGAGGAGGAGGGGCTCTCCAGGGAATGTGTCATATGACACACGG	837
OY	762	TTGGCAATTCAGTCGAGACGATTCGGATTCCTATCAC---CAAGAAGAGATTAACCAACAA	818
Db	838	GTCGCAAGCCATTTGTCACGACGAGGAGACCCCAATACGAGACCAAGAAGGTATGAAGCAAC	897
OY	819	CGTCCCGCGGACGCAAAATGCCCTGCCGTCCTTTTGAATTTCCAAGAGAAATTAACAAATTCG	878
Db	898	CTCCCGACGGGAGCAAAATGCCATGACAGGTGTGGTGAAGTTGGAGGAGAAATTAACAAATTCAG	957

QY	879	TGCTACGTCAATGCCAAGCCACCGGCCGCCGAGCGGCGCAGAAAGGATATACAGCGTT	938
Db	958	ACAGATTATGTAGTGAAGAAGATTCGTG-----AGAAAGCATATGGGTGCCCTT	10020
QY	939	TATAGATGACACTCAAAAGAGAGATTAAAGACTGTGGAGCATGTTTATGTCTTGGCATGCTTT	998
Db	1003	TGTTAGGGACTTGAAGAACAGTTTATAGACCGCTGGAGCAGAGTGTATGTGTGGCACCGCGCT	10678
QY	999	GTTGTGATATTTGGGGTGGCCCTTGCGCCGCGAGTGGCTTGCCCTGAGGACAGCTGTGAT	10588
Db	1063	TTGTGGGATATGGGGTGGGGGTTCAGACCCAAAGTTCGCGGCGCATGCCCCAGCGTTAAGTTGT	112228
QY	1059	TCACCCAGTGGCTTTCACCAGGCGCTGCAGATGACAGATGGAGAGATTGGCGGTGGATAAAGT	11188
Db	1123	CACGTCCAGAGCTGTCCCATATGGACTTAAATTTGAATAAGAAAGATTTTAGCGGTGGATTAAGAT	118288
QY	1119	TGTTCTTCATTAAGGTGCGGCGTGGCGCCGCGGAGAAAGGCTGAGAGAGATTTACGAAGACT	11788
Db	1183	CGTCAATGACGGAATTTGGACTGTGTGCCACCACACCTGGCTCACCTTTTGTACGAGGGCGCT	12428
QY	1179	TCATGCTCATTTGGAAAAAGTTGGGATTCGACGGCTGTAAAGATTTGACGTTATCCACTATT	12388
Db	1243	CCACTCCCGGTTTGGAAATCTCGGGGTATTTGACGGTGTAAAGTTTGAAGTTATACACTTCT	130202
QY	1229	GGAATATGTTGTGTGAAGACTATGGAGGAGAGTGGATTGGCAAAAGCATATTTACAAAGC	123988
Db	1303	CGAATATCTATTCGAGAGAAATACGGTGGCGGTGTGGAGCTAGCCAAAGCTTATTTACAAAGC	136288
QY	1299	AATGACCAAAATCAATTAATTAACATTTTAAAGAAATGGAGTCATTTGCAAGATATGACAA	135888
Db	1363	GCTACATGCTTCGGTGAAGAAAGCATTTTCAAAAGGCATATGGGGTCAATTTGCGAGCATGGACAA	142222
QY	1359	TTTGTAACGACTCATGTTCTCTTGGACGAGAGCATCTCTCTGGTGTGTTGATGATGA	141888
Db	1423	TTGTAAATGACTTCTTCTCTCTGTTGATACCGAAGCATAGCCCTTGGGGCGGTAGAGATGA	148288
QY	1419	CTTTTGGTGCACGAGACCCCTCTGTGATATCCAAACGATACGTTTGGCTTCACAGATGTCA	147888
Db	1483	TTTTTGGTGCACATGATCCCTCTGTGAAGATCCAAATGGCACGATTTGGCTTCACAAAGGCTTCA	154288
QY	1479	CATGCTTCATTTGTGCCCAACACAGCTTGTGGATGGGGAACCTTCAATCCACCCGACTGTGGGA	153888
Db	1543	CATGCTGCACCTGTGCTTCAACAAAGCTTTGTGATGGGGAATTTTATTTACCGCGGATTTGGGA	160288
QY	1539	TATGTTCCAAATCCACCCACCCCTTGTGCGCGCTTCCATGCTGCTCTCGAAGCCATCTGCG	159888
Db	1603	CATGTTCCAGTCCACATCACCCCTTGTGCCGAATTCATGACGCGCTTAAGGCGCATCTGCG	166288
QY	1599	TGGCGCCATCTATGTATGTATCTGTGGGAAAGCATAACTTTGATCTTTCTGAAAAAACT	165888
Db	1663	TGGACCAAGTTTACGTTAAGTATGTTATGTTTGGAAAGCAACATTTCAAGTTGCTCTCAAGAGCT	172222
QY	1659	AGTCTCTTCGATGGATGCATCTCTGGAAGTGAATGTGATGACATCCGACCGCGAGGATG	171888
Db	1723	CGCTTGGCTGTGATGGAGCAATTTTGGCTTGTCACTATGTGACTTCCACACGAGACTG	178288
QY	1719	TTTGTTTGAAGACCCCTTGTGCATTAATGAGAGAAACTATGCTTAAAGTTTGGAAATCTCAACA	177888
Db	1783	TTTGTTTGAAGACCCCTGTGATGATGGGAAGCAATGCTCAAAATTTTGGAAATCTCAACA	184288
QY	1779	GTTTACTGTGAGTATGTGTGCTCAATTCAGTCCAGAGGAGGAGTGTGTCTGTGAGACAGC	183888
Db	1843	ATATACAGGTTTGTGGCTTATTTAATTTGCAAGAGAGTGGGTGTCTCCCGTAATCAG	190202
QY	1839	CCGGAACCATATGCTTTTCAACATATCTCAAAACGAGTGAATCTCAAACTAACCCAAAAAG	189888
Db	1903	GAGAAACAAAGTCCCTGTGAATTTTTCACAAACTGTGACATCTTAAAGGATCTCTCAAGA	196288
QY	1899	CATGATGATGCACAGTGGAGAAAAACCTTCTCTTATGGAAGCGTTAAACCTTTGGCGCT	195888
Db	1963	CATTGATATGAGCAATATGGGAAAAAGCCCAATATGCAATAAAAGGGAATGAATGTGTTGCTGT	202288
QY	1959	TTACCTCTATCAAGCCAAAAAACTTATCTCTCAAGGCCCTCTCAAGATCTTGCATATGCT	201888

Db	2023	ATATTGTTCAAGGACCAACAACTAAGCTCATGAAAGCATCAGAGAAATTGGAAGCTTTC	2082
QY	2019	TTCTTACCATTGCAATTCGAGCTCATCTACTGTTTCCACAGGTGACCAACATTCACAAAC	2078
Db	2083	ACTTAGAGCATTTACTTTTGTAGACTTTTGAAGAGTGTCTCTCATGATTTGCTGTCAAAAAA	2142
QY	2079	TTCTCTACACTTTTGGCCCAATTTGGGCTGGTGAACATGCTTAAACACTAGTGGAGCCATCCA	2138
Db	2143	GTTAATTCATATTTGCTCCCAATGTGATTTAGTGAACATGCTTAAACACTGTGTGTGCCATTTCA	2202
QY	2139	ATCTGTGACACTTGTGACGATGACCTAAGCTCAGTCAGATTTGGTGTCAAAAGGCTGTGGGA	2198
Db	2203	GTCACATGGAGTTTGGACAACACATAGATGATGCTCAAAATTTGGGTTAGGGTGTGGGGA	2262
QY	2199	GATGCGAGATTTTTCATCGAAGAAAACCAAGGCTTGTGCTATGATGCGAGAGATTTGG	2258
Db	2263	GATGGAAGGTTTGGATCATAGAGAAACCAAGTTAGTTGCCAACTAGATGGGCTAGTTGTAAA	2322
QY	2259	GTTCAAGTATGATCAGCAACAAATGGTGGTTCACATGCGCATGGCCAAATTTGATCTTC	2318
Db	2323	ATTGATTTATG---AGATATAAATTCGTGAGATGCAAGTATCCCTAGTGGCTTCAAA	2379
QY	2319	ATCG 2322	
Db	2380	ATG 2383	
RESULT 6			
LOCUS	E37133	2690 bp	DNA
DEFINITION	Ratifinose synthase gene.		
ACCESSION	E37133		
VERSION	E37133.1 GI:18626585		
KEYWORDS	JP 2000166562-A/1.		
SOURCE	unidentified.		
ORGANISM	unidentified.		
REFERENCE	1 (bases 1 to 2690)		
AUTHORS	Watanabe, E. and Oeda, K.		
TITLE	Ratifinose synthase gene		
JOURNAL	Patent: JP 2000166562-A 1 20-JUN-2000;		
COMMENT	SUMITOMO CHEM CO LTD		
	OS Beta vulgaris L.		
	PN JP 2000166562-A/1		
	PD 20-JUN-2000		
	PE 04-DEC-1998 JP 1998345590		
	PR		
	PI EIJIRO WATANABE, KENJI OEDA		
	PC C12M15/09, C12N1/21, C12N5/10, C12N9/88, C13Q1/68//C12M15/09, PC		
	C12R1/91,		
	PC (C12N9/88, C12R1/91), C12M15/00, C12M5/00, (C12M15/00, C12R1/91) CC		
FEATURES			
source	FT key Location/Qualifiers		
	1..2690		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	762 a 507 c 641 g 778 t 2 others		
ORIGIN			
Query Match	41.5%; Score 975.4; DB 6; Length 2690;		
Best Local Similarity	67.8%; Pred. No. 6.8e-280;		
Matches 1443; Conservative 1; Mismatches 662; Indels 21; Gaps 5;			
QY	183	GGTTTCGGTTGGTCTTTGTTGATTCGACGCGTCGGAACCTGATAGCCGACATGTTGT	242
Db	403	GGGTTTTCATGGGCTTTTGTGGGTTTTCGATGCTCCAGAGCCCAAGGCCCGCACGTTGT	462
QY	243	TTTCGATTTGGGAAGCTGAAGCATATTCGGTTTATGATGATTTTCAAGTTTAAAGTTTGGTG	302
Db	463	ATCCGTGGGCGAGCTCAAGGAGAAATTCCTTCATGATATCTTCAGGTTTCAAGGATGATGGT	522

Qy	303	GACTACACACTGGGTTGTCGAAATGTTGGGATCTTGTAATCGAGACTACGATTTGGAT	362
Db	523	GACTAACCCATTGGAGCTGGGTCACMAATGGGGGGACCTTGGAGCTAGAGCCAAATTTCTAT	582
Qy	363	CCTTGAGAACTCAGA-----TTCTGTGTGACCCGATATGTTTTCTTCTTCCGATCGTTGA	416
Db	583	CCTTGATTAACCTCAGATGAAAGGTTTTGGGCGCGTCCCTATATTTGATCCCTCCATTGATCGA	642
Qy	417	GGGACGCTTCCGAACTCGATTTAGCCTGG-- -ATGATGACTTTGTCGATGTTTTGT	473
Db	643	AGGCCCCATTGGGGACTCTCTCCAGCCGGGTTCTGTATGACTATGTGCAATATGTGT	702
Qy	474	CGACAGTGTGTTCTCGAAAGTTGGTTGATGCATCGTTCGAACTATGTTGTATCTTCAATGC	533
Db	703	TGAGAGTGGGTCCACTAAAGTTGTGCGAGACHTGTTCCGGGCTGTTCTTTATATACGGGC	762
Qy	534	TGTGTGATGATCCGTTTGCACTTTTAAAGAGCGATGAAAGATCGTGAAGACCCATCTTGG	593
Db	763	TGGGCCCTGACCCATTAAAGTTAAATTAATTAATGAAAGAAAGTCAAGGCCATTAAAG	822
Qy	594	AACTTTTCGGTTGTGGAGAGAAAGCTCACAGGATCGTGGCAAAATTCGGTGGTG	653
Db	823	GACTTTTAAACTCTTGATGATACAAACCTCTCCAGGATATGTGGCAAGTTTGGATGGTG	882
Qy	654	CACGTGGAGCGGTTTATCTTAACGAGTTTCATCCACAGGCGCTTAATAGAAAGCGTGAGCA	713
Db	883	TACATGGGATGCAATTTTAACTCAAAATNADAGCCMTATGTTGTGGGAAGAGTTAAAG	942
Qy	714	TCCTCTGCAGCGGCTTTCCTCCGGTTTTAGTCTTAATCGACGATGTTGGCAATCCAT	773
Db	943	ACTGTGCAAAAGGGGGTCCACCGGCTCGTACTCATTTGATGATGGTGGCAATCTAT	1002
Qy	774	CGGACGAGATTGGATGCCATCCATCAACA-- -AAGAAGAAATGACCAAAACCGTCCCGGGA	830
Db	1003	TTTGATATGACGATGATCTCCGATTCGATCCCAAGAAAGGAGATAAACCGGACTTCTCGCGGGA	1062
Qy	831	GCAAAATCCCTGGCGCTTTTGAAATTCCAAGAGAAATTCAAAATTCCTGTACTACGTCAA	890
Db	1063	GCAAAATGCCATGTATGATGTATCAAGTACGAGGAAACCTCAAGTTTAAAGGACTATAAAG	1122
Qy	891	TCCCAAGGCCACGGGCCCCGAG-- -CGGCGCAAGAGGGGATGAAGGCGTTTATATGATGA	947
Db	1123	CCCAAAATTAATGGGCGCATAGAGTATCCCAATATGGGAAAGAGCGGCTTTGTTAGGA	1182
Qy	948	ACTCAAAAGAGAGTTTAAAGCTGTGGAGATGTTATGTTGGCATGCTTTGTGTGATA	1007
Db	1183	CTTTAAAGAGAGTTTCAAACTCTTTABGATGTGTTATGTTTGGCATGCTTTTACGGGCTA	1242
Qy	1008	TTTGGGTGGCCTTTCGCCCGCAGGTGCTGGCTGGCTGACCTGTGATTCAGCCAGT	1067
Db	1243	TTTGGGAGGGGTTAAGGCCCAATTTCCAGGCGTACCRHAGGCCCAAGTATACCCCAA	1302
Qy	1068	GCTTTTACACAGGGCTGCAGATGACGATGAGAGATTTGGCGGTGATTAAGATTTGTTCTTA	1127
Db	1303	GCTTTTCCCCGGGCTTGAGATGACATGGAAGATCTACGTGGTAAATTTGTTAATA	1362
Qy	1128	TAAAGTGGGCTGTGTCGCCCGGGAAGAGCTGAGAGATGTACGAAGACTTCACTGTCA	1187
Db	1363	TGTGATTTGGCTTGTCCAGGCTCTATTAAGGCCCAAGAACTTTTGAAGGGTTGTCATTTCA	1422
Qy	1188	TTTGGAAAAAGTTTGGATTCACGCGTGTAAAGATTTGACGTTATTCACCAATTTGAGAGATGT	1247
Db	1423	TTTGGAAAAATTTGTGGATTTATGTGAAGTCAAAAGTTGTTCATCTCAATTTGTTGTGAGATAT	1482
Qy	1248	GTGTGAAGACTATGACGAGGAGAGTGTATTTGCCAAAGCATATTCAAAAGCATGACCA	1307
Db	1483	GGCAGAGAGACTATGAGAGAGAGATGTAACTAGCAAAACATCTCTTAAGCAATAACAGA	1542
Qy	1308	ATCAATTAATTAACATTTTAAAGAAATGAGTCAATTCGAAGTATGGAACATTTGTAACGA	1367
Db	1543	ATTCAGTGCATTAACGATTTTCAAAAGCAACGCTGTGATTTGTACATGTAGGACAGTGCMAA	1602

QY	1368	CTTCATGTTCTCTTG6GACGGAAACATCTCTCTTGCTGCTGTTGGTATGACTTTTGGTG	1427
Db	1603	TTTCTATGCTCTCTTG6TACTAGACCAATTTGCTCTTGCTGCTGTTGGGATGACTTTTGGCC	1662
QY	1428	CACGAGACCCCTCTGTATCTCAAAACGATCGTTTTGGCTCCAAAGATGTCATCATGTTTCA	1487
Db	1663	AACGTATCCGCTGTGAGATATTAAATGTGACATATTGGCTCCAAAGCTGTCAATTGTGTCA	1722
QY	1488	TTTGGCCAAAGACAGCTTGTGGATGGGGAATTCATCCACCCTGACTGGGATATTGTCA	1547
Db	1723	TTTGGCTCAAAATAGCTTATTGATGGGAAACCTTATACACCCTGACTGGGACATGTTTCCA	1782
QY	1548	ATCCACCACCCCTGTGTGCCCCCTTCATGCTGCTGTGAGGCATCTCTGGTGGCCGAT	1607
Db	1783	ATTCACACACCCCTGTCTCAATTTATGCTGCTATCTGTGGATTTTCTGTGTGGACCAAT	1842
QY	1608	CTATGTTAGTATCTGTGGGAAAGCATACTTGTATCTTCTGAAAACCAATAGTGCCTCC	1667
Db	1843	TTATGTTAGTATGTTTGTTTGGCAACGATACATCCCCCTTGCTCAAAAGGCTGTCTTG6C	1902
QY	1668	TGATGATCGATCTCTTGGAAAGTAGTACTATGACACTCCCGACTCCGATTTGTTGTTTGA	1727
Db	1903	TGATGATCGATCTCTTGGCTGTGGAGTAGACATGACTCTCTACTAAGATGTGCTATTGTT	1962
QY	1728	AGACCCCTTTCATATGAGACAACATGCTTTAAGATTTTGAATCTCAACACGTTACTGG	1787
Db	1963	AGATCTCTTTCGACATGTCGAAACCAATGTCTCAAAATTTTGGAACTCAACACGATCAATGG	2022
QY	1788	AGTATGTTGTCATTCAAACGCCAAGAGAGATGTCGTGTGAGACACGCCGCAACCA	1847
Db	2023	AGTGTGTTGACTCTTCAATTGGCCAAAGGAGGAGTGGACCCGTGAGTCTCGAAAAATCT	2082
QY	1848	ATGCTTTTTCACAAATCTCAAAACGAGTGACATCCAAACCTAACCCMAAGACATAGATG	1907
Db	2083	ATGTTTTCACAGATATTCAAAACCTATTTCCTGTCGAAGACATGCCAAAGATGTTGATG	2142
QY	1908	GCACAGTGTGAGAAAACCTTATCTGTATTGAGGCGTTTAAACCTTTGCGCTTTTACCTCTA	1967
Db	2143	GGAGAACGGACACAAAGCATTTCCCATTCAAAGGAGTGGATTTTGTGCATGTACTCTAC	2202
QY	1968	TCAACCCCAAAAACCTTATTCCTCTCCAAAGCCCTCTCAAGATCTTGTACATAGCTCTTGACC	2027
Db	2203	CAGGAAAAAAAGCTAATCTCTTCACACATGTGACACCATTTGAATATATCTACTGTATGC	2262
QY	2028	ATTGCAATTCGAGCTCATCACTGTTTCACAGTGAACCAACTCATCAACCTTCTTACA	2087
Db	2263	CTTGCAATTACGAGCTTATTGTAAGTCTCTCCGATGACAAATTCTACCCGTGGAGTGCATGC	2322
QY	2088	CTTTTGGCCCAATTGTGGCTGGTGAACATGCTTAACTAGTGAAGCCATCTCATCTGTGGA	2147
Db	2323	ATTTGTCCACCCATTAAGATTAGTAAACATGTCTCAACGCCGAGGGGAGTCAAGTCTTTTGG	2382
QY	2148	C-----TATACGATAGCACTTAAGCTCAGTCGAGATTGGTGTCAAAAGGATGGTGAAT	2201
Db	2383	CATCACTGATAGATTAATAGCATTAAGATGTTTTCAGGTTTGTATTAAAGGGGCCGAGAAAT	2442
QY	2202	GGGACTATTTCATCGAAAAACCAAGGCGTTGCTGATATTGATGGGAGGATGTTGGGTT	2261
Db	2443	GATGGTTTATTCTACAGAAAAAGCCAAAGCGTGTAGATTAATGGAAGAAACATGGAATT	2502
QY	2262	CAAGTATGATAGGACCAATGCTGCT	2288
Db	2503	TGAGTATGAAAGAGCATGATTTAAGCT	2529

RESULT 7			
BD010263	BD010263	2343 bp	DNA
LOCUS			Linear
DEFINITION	Mutant protein, its DNA and use thereof.		
ACCESSION	BD010263		
VERSION	BD010263.1	GI:18638636	
KEYWORDS	JP 2001078783-A/1.		
SOURCE	Glycine max.		
			PAT 31-JAN-2002

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
ORGANISM	Glycine max			
	Eukaryota, Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.			
	1 (bases 1 to 2343)			
	Watanabe, E. and Oeda, K.			
	Mutant protein, its DNA and use thereof			
	Patent: JP 2001078783-A 1 27-MAR-2001;			
	SUMITOMO CHEMICAL CO LTD			
	OS Glycine max (soybean)			
	PN JP 2001078783-A/1			
	PD 27-MAR-2001			
	PF 03-JUL-2000 JP 2000200571			
	PR			
	PI EIJIRO WATANABE, KENJI OEDA			
	PC C12N15/09, A01H5/00, C12N5/10, C12N9/10, C12Q1/68, C12N15/00, PC C12N5/00			
FEATURES				
source	1..2343	Location/Qualifiers		
	FT	(1)..(2343).		
BASE COUNT	597 a 493 c 648 g 605 t			
ORIGIN				
Query Match	41.1%; Score 966.8; DB 6; Length 2343;			
Best Local Similarity	66.0%; Pred. No. 2,5e-277;			
Matches 1508; Conservative	0; Mismatches 737; Indels 39; Gaps 6;			
OY	57	TCACATGTCGTACCGTTTGCATCGACGATGGATTTCTACTGTGAACGGTCATTCGTT	116	
Db	57	TAAATTGCCCTTGTCCATAAACCCCTGAAGAGATCAATTTCTCGCCAAAGCCACCCCTTT	116	
OY	117	TCTGTCCATGTTCTCTGAGAACATGTTGCTTCTCTCTCTCCGTAC-----AC	164	
Db	117	TCTCAGGAAAGTTCGGAACATATATATGTACCCCTTCACCCATCGACGCCAAGATAG	176	
OY	165	TTTCGATAGACAAGTCCCGGTTTCGGTTGGTTGCTTTGTTGGATTGCGACGCTCGAAC	224	
Db	177	TAAAGAACACAGGACGACGACGACGACGTCGTGATGTTGCTTCTCGGCTTCACGCGGACGAC	236	
OY	225	TGATAGCCGACATGTTGTTTCGATTGGGAAGCTGAAGATATTCGGTTTATGATATTTT	284	
Db	237	CAGAACCCGACACGTGGCTTCCCTGGGGAAGCTCAGAGGAATAAATTCATGACATATTT	296	
OY	285	CAGTTTAAAGTTTGGTGGACTACACACTGGTGTGGTGAATGTTGGGGATCTTGAATC	344	
Db	297	CCGTTTAAAGTGTGTGGACACCACTACATCGGTGGTATGCAACGACACGAACCTGGACCA	356	
OY	345	GGAGACTAGATTGGATTCCTTGAAGAACTGAC---ATTCTGGTCGACGATATGTTTCT	401	
Db	357	CGAGACACAGATGATGATCTTCTTCGACAAAAAGACACACAGCTGGACCCCTTTGTTGAT	416	
OY	402	TCTTCGATTCGTTGAGGACCGTTCCGAACCTCGATTAGCCCTGGGGATGATGATTTGT	461	
Db	417	TCTCCGATCTCCCAAGCTCGTTCGAGACCTCCCTGCAACCCGGTTGGATGATTACGT	476	
OY	462	CGATTTTGTGTCGAGATGTTGTCGAAAGTTTGTATGATGATGTTCCCAAGATATGTT	521	
Db	477	GGACGTTTGCATGGAGGACGGCTCCACACAGTCTCTGAGTCCAGCTTCGGGAGGTGCTT	536	
OY	522	GTATTTTATGCTGTGATGATCCGTTTGACCTTTTAAAGGGGAGTAAAGATCGTAG	581	
Db	537	ATAGCTCCACGTTGGCCATGACCCCTATACAGTTGCTTGAAGAGCAACTAAAGTGTAG	596	
OY	582	GACCAATCTTGAACATTTTTCGTTTGTGGAGAGAGAAGATCCACAGGATATCTGACAA	641	
Db	597	GATGATTTTGGGGAATTTCAAGCTTCTCGAGAGAGAAAACCGGCCACGATGCTGACAA	656	
OY	642	ATTGCGTTGGTGCACGTGGGACGCGTTTACCTTAACGGTTTCATCCACAGGGCGTAAAGA	701	

Db	657	GTTTGGTGGTGATGACATGGGACGGGTTTAACTTGAAGGTGCATCCCTCAGGTGTGGGA	716
Qy	702	AGCGCTGGGCATCTCGTCGACGGCGGTTATCTCCGGTTAGTCTAATGACGATGG	761
Db	717	AAGGGTGAAGAAGGTTGGTGGAGGAGGAGGTCCCTCCAGAGGATGTCTAATGACAGACGG	776
Qy	762	TTGGCAATTCATCGGACACGATTCGGATCCATCAC--CAAGAAGATGACCAAC	818
Db	777	GTGGCAACCCATTGTCTGCACGACGACCCCATTAACGGACCAAGAGGATGAACCGAAC	836
Qy	819	CGTGGCGGCGACGAATTGCCTGTCCGCTCTTTGAATTCCAAAGAAATTACAATTTCCG	878
Db	837	CTCCGCGAGGGAGGAATATGCCATGACAGTTGGTGAAGTTGGAGGAAATTAACAAGTTAG	896
Qy	879	TGCATACGTCATCCCAAGGCGACCGGCCCGGACGAGGAAGGGATGAAGCGCTT	938
Db	897	ACAGTATTTGTGTGAAGGATTTCTG-----AGAAAGTATGGGTGCTT	941
Qy	939	TATAGATGAACATCAAGAAGATTTAAAGCTGTGAGACATGTTTATGTTTGGCATGCTTT	998
Db	942	TGTTAGGACACTTGAAGAACAAGTTTAGACCTGTGAGACAGGTGTATGT--GCACGCGCT	998
Qy	999	GTCGTGATATTGGGTGGCGCTTCGCCCGCAGGTGCGCTGCGCTCCGAGGACAGTGTAT	1058
Db	999	TTGTGGGATTTGGGGTGGGGTCACAGCCCAAGTTTCGGGGATATGCCCAAGCTTAAGGTGT	1058
Qy	1059	TCAGCCAGTGTCTTCCACAGGGCTGCAGATGACGATGAGAGATTTGGCGGTGATAGAT	1118
Db	1059	CACCTCCGAAGCTGTCCAAATGAGACTAAATTTAGCAATGMAAGATTTAGCGGTGATACAT	1118
Qy	1119	TGTTCTTCATTAAGTTCGGGCTGTCCCGCCGGAAGAGCTGAGAGATGTACGAAGACT	1178
Db	1119	CGTCAGTAAAGGATTTGGAGCTGGAGCTGCACACCTGCTCACCTTTTGTATGAGAGGGCT	1178
Qy	1179	TCATGCTCATTTGGAAAAAGTTGGGAGATGACAGCGGTGAAGATTGAGCTATCCACTCAT	1238
Db	1179	CCACTCCCGTTGGAAATCTGGGGTATGTACGGTGTGAAGTTGACGTTATACACTTCT	1238
Qy	1239	GGAGATGTTGTGTAAACACTATGAGAGGAAGTGTGATTTGGCAAAAGCATATTACAAAG	1298
Db	1239	CGAATGCTATCCGAGGAATTAAGGTGGCGCTGTGAGCTAGCAACAAGCTTATTACAAAGC	1298
Qy	1299	AATGACCAATCAATTAATAAATCAATTTTAAAGSAAATGAGTCAATGTCAGATGGAACA	1358
Db	1299	GCTCACTGCTTCGGTGAAGAAGACTTTCAAGGCAATGGGGTCAATTTGCGAGCATGAGACA	1358
Qy	1359	TTTGAACGACTTATGTTCTTGGACAGCAAGCTATCTCTTGGTCTGTGGTGTGATGA	1418
Db	1359	TTTGAATACCTTCTTCTCTCTGGTACCGAAGCATATAGCCCTTGGCGGTGAGAGATGA	1418
Qy	1419	CTTTTGGTGCAGGACCCCTCTGGTGTCAAAAGGCTAGCTTTTGGCTTCAAGAGATGCA	1478
Db	1419	TTTTTGGTGCAGTATGATCCCTCTGGAGATCAAAATGACAGTATTTGGCTTCAAGGCTGCA	1478
Qy	1479	CATGCTTCATTTGTCCCAAGCAAGCTTGTGATGGGGAATTCATCCACCTGACTGCGGA	1538
Db	1479	CATGGTGCAGTGTCTTCAACAACAGCTTGTGATGGGGAATTTATTACGCGGATTTGGGA	1538
Qy	1539	TATGTTCCAATCCACCAACCCCTGTGGCGCTTCCATGCTGCTCGTCCAGGCATCTCGG	1598
Db	1539	CATGTTCACAGTCACTACCCCTTGTGCGCAATTCATGAGAGCTCTAGGGCCATCTCGG	1598
Qy	1599	TGGCGCCGATATTTAGTATCTCGGGGAAGCATTAATCTTGATCTTGTGAANAATCT	1658
Db	1599	TGGACCAATTTAGCTTATGTATTTGTGTGGAAAGCAACACTTCAAGTTGTCTAGAGGCT	1658
Qy	1659	AGTGCCTTCGATGATGATCGATCTCTCGAAGTGAATCTATGACATCCGAGCTCGGATGT	1718
Db	1659	CGCTTTGGCTGATGGAGCATTTTGGGTGTCAACCATATGACATCCCGACGAGACTG	1718
Qy	1719	TTTGTTTGAAGACCTTTGCATATATGGAGAACTATGCTTAAGATTTGGATCTCAACA	1778

Db 958 ACAGTATTGTAGTGGAAAGATTCTG-----AGAAAGGTATGGGTGCTT 1002
 QY 939 TATAGATGAACCTCAAGAGAGATTAAAGACTGTGAGACATGTTATGTTGGCATCTTT 998
 Db 1003 TGTTAGGACCTTAAGGAACAGTTTAGAGACAGGTGTATGTGTGGACCGGCT 1062
 QY 999 GTGTGATATTTGGGGTGGCCCTTCGCCGACAGGTGCTTGCTGAGGACAGTGTGAT 1058
 Db 1063 TTGTGGGTATTTGGGGTGGGTGACACCCAAAGGTTCCGGCGCATCCCAAGGCTAAGGTGT 1122
 QY 1059 TCAGCGAGGCTTTACACAGGCTGCAGATGAGATGAGAGATTTGGCGGTGATTAAGT 1118
 Db 1123 CACTCCGAAAGCTTCCAAATGAGCTAAATTAATGCAATGAGGATTTAGCGGTGATTAAGT 1182
 QY 1119 TGTTCCTTCAATAGGTGCGGTGCTCCCGCGAGAAAGGCTGAGAGATGATGACAAAGACT 1178
 Db 1183 CGTCAGTAAACGAGATTGGACTGTGTGCACACACCTGGCTCACCTTTTGTAGCAGGGGCT 1242
 QY 1179 TCATGCTCATTTGGAAAAAGTTGGATTCGACGGTGTTAAGATTGACGTTATCCACTATT 1238
 Db 1243 CCACCTCCGTTTGGAAATCTGCGGGTATTGACGGTGTAAAGTTAGAGTTATACACTTGGCT 1302
 QY 1239 GGAGATGTTGTGGAAGACTATGAGAGGAGAGTGGATTTGGCAAAGGCATATACAAAGC 1298
 Db 1303 CGAGATGCTATCCGAGGAATACGGTGGCGGTGTGAGCTAGCCAAAGCTTATTACAAAGC 1362
 QY 1299 AATGACCAATCAATTAATTAACATTTTAAAGGAATGAGATCATGCAAGTATGGAACA 1358
 Db 1363 GCCTACTGCTGTGGTGAAGAAGACTTTCAAAAGGCATGGGGTATTGGCAGCATGGAGCA 1422
 QY 1359 TTGTAAACACTTCATGTTCTCTTGGACACGGAAGCTATCTCTTGTGTGTGTGTGATGA 1418
 Db 1423 TTGTATGACTTCTTCTCTCTGTGTACCGAAGCATAGCCCTTGGCGGTAGAGATGA 1482
 QY 1419 CTTTGTGTGACGAGCCGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1478
 Db 1483 TTTTGTGTGACTGATCCCTGTGAGATCCAAATGACAGTATGTGCTCCAAAGGTGTCA 1542
 QY 1479 CATGTTTCATTTGTGCAACGACAGCTTGTGATGGGAACTTCATCCACCTACTGGGA 1538
 Db 1543 CATGGTGAAGCTGTGCTTACAAACAGCTTGTGATGGGAAATTTTATTCAGCGGATTTGGGA 1602
 QY 1539 TATGTTCAATCCACCACCTTGTGCGGCTTCATGCTGCTTCGAGCCATCTCTGTG 1598
 Db 1603 CATGTTCCAGTCCACACCTTGTGCGCAATTCATGCG- GCCTCTAGGGCCATCTCTGTG 1661
 QY 1599 TGGCCGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1658
 Db 1662 TGGACCACTTTACGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1721
 QY 1659 AGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1718
 Db 1722 CGTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1781
 QY 1719 TTTTGTGGAAGCCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1778
 Db 1782 TTTGTTTGAAGCCCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1841
 QY 1779 GTTCACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1838
 Db 1842 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1901
 QY 1839 CGCAACCAATGCTTTTCAACAATACTCAAAAGAGTGAATCCAAACTAACCCAAAGA 1898
 Db 1902 GAGAAACAGAGTGGCTCTGAATTTTCAACAATGATGATGATGATGATGATGATGATGATGAT 1961
 QY 1899 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1958
 Db 1962 CATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2021
 QY 1959 TTACCTCATGCAAGCAAAACTTATCTCTCCAAAGCCCTCTCAAGATCTTGACATAGC 2018
 Db 2022 ATATTTTGTTCAGGACCAAACTAAAGCTCATGAAAGGATCAGAGAAATTTGGAAGTTTC 2081

QY 2019 TCTTGACCATTCGAATTCGAGCTCATCTGTTTGCACGATGACCAACTCATCCAAC 2078
 Db 2082 ACTTGAGCCATTTACTTTTGTAGCTATTGACAGTGTCTCCAGTATGTTGTGTCAAAAAA 2141
 QY 2079 TTCTCTACACTTTTCCCAATTTGGCTGGTGAACATGCTTTAACACTATGAGACCATCA 2138
 Db 2142 GTTAATTCATTTTGTCCCAATTTGATGATGAAACATGCTTAAACCTGTGTGTGTCA 2201
 QY 2139 ATCTGTGACTATGACGATACCTTAAGCTCATGCGAGATGATGATGATGATGATGATGATGAT 2198
 Db 2202 GTCTCATGAGATTTGACACACCATAGATGTGTCTCAAAATTTGGGTTTAGGGGTGGA 2261
 QY 2199 GATGCGATATTTGATTCGAATAAACCAAGGCTGTGCTATGATGATGATGATGATGATGAT 2258
 Db 2262 GATGAAGGTGTTTGCATCAAGAAACACAGTATGATGCAAACTAGATGGGTACTTGTAA 2321
 QY 2259 GTTCAAGTATGATCAGACCAAAATGCTGTGTTCAAGTCCATGCGCAATTTATCTTC 2318
 Db 2322 ATTTGATTTATG--AGGATTAATGCTGAGAGTGCACAAATTCCTGGCTAGTGTCTCAA 2378
 QY 2319 ATCG 2322
 Db 2379 ATTG 2382

RESULT 9
 PSAA26475 2671 bp mRNA linear PLN 15-JAN-2002
 LOCUS PSAA26475
 DEFINITION Pisum sativum mRNA for raffinose synthase (rfs gene).
 ACCESSION AJ426475
 VERSION AJ426475.2 GI:18181864
 KEYWORDS raffinose synthase; rfs gene.
 SOURCE pea.
 ORGANISM Pisum sativum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae;
 Pisum.

REFERENCE 1
 AUTHORS Peterbauer, T., Mach, L., Mucha, J. and Richter, A.
 TITLE Molecular characterization of raffinose synthase from pea (Pisum sativum L.) seeds
 JOURNAL unpublished
 REFERENCE 2
 AUTHORS Peterbauer, T.
 TITLE Direct Submission
 JOURNAL Submitted (02-JAN-2002) Peterbauer T., Institute of Ecology, University of Vienna, Althanstrasse 14, Vienna, A-1090, AUSTRIA
 REFERENCE 3
 AUTHORS Peterbauer, T.
 TITLE Direct Submission
 JOURNAL Submitted (09-JAN-2002) Peterbauer T., Institute of Ecology, University of Vienna, Althanstrasse 14, Vienna, A-1090, AUSTRIA
 COMMENT On Jan 16, 2002 this sequence version replaced gi:18071124.
 FEATURES
 source

1..2671
 /organism="Pisum sativum"
 /cultivar="Wunder von Kelvedon"
 /db_xref="taxon:3888"
 /tissue_type="seeds"
 /country="Austria"
 148..2544
 /gene="rfs"
 148..2544
 /gene="rfs"
 /EC_number="2.4.1.82"
 /function="galactinol:sucrose galactosyltransferase"
 /codon_start=1
 /evidence="experimental"
 /product="raffinose synthase"
 /protein_id="CAD20127.2"

[illegible]

Dd	1105	GGTATTAATGAGGAAGAAAGGTTTGTTGGTTTGTGAGGGATTGTAGAGACGATTT	1154
Qy	964	AAGACTGTGAGCATGTTTATGTTTGGCATGCTGTTGTGTGGATTTATGGGGTGCCCTTCG	1023G
Dd	1165	AGGAGTGTTCAGAGTGTTTATGTCTGGCATGCGCTTTGTGGGTATTGGGGTGGGTTTGA	1224A
Qy	1024	CCGAGGTGCCCTGGCTTTCCTGAGGCACGTGTGATTTCAGCCAGTGCCTTTCACCAAGGCTG	1083G
Dd	1225	CTTAAGTGTGTGGAAATGCGCGAAGCTTAAGGTGTGTTTCCGAACCTCTTCGCGGGCTG	1284A
Qy	1084	CAGATGACATGAGGAGATTTGGGGGGGAGTAAGATTTGTTCTTCATTAAGAGTGCGCGTGC	1143G
Dd	1285	AAGATGACGATGAGGAGTTTATGGGAGTGATTAAGATTTGTGAGAAATGTTGTGGGCTACTG	1344A
Qy	1144	CCGCGGAGAGAGCGCTGAGAGATGTACGAAGACTTCATGCTCATTTGGAAAAATCTGGG	1203G
Dd	1345	CTCTCAATTTTGGCCGAGGAGATGTTGATGGGATTTCACTCTCATTTTGGAGTGGCGGGA	1404A
Qy	1204	ATGAGCGGTGTTAAGATTGACGTTATCCACTATTGGAGATGTTGTGGAAGACTTATGGA	1263G
Dd	1405	ATTACCGTGTAAAGTTGACGTTATCCATTTCCTTGAGTPTACTATCAGAGCAATATGCT	1484A
Qy	1264	GGGAGACTGATTTTGGCAAGGCAATTTTACMAAGCAATACCAATTCATTAATTAACAT	1323G
Dd	1465	GGAGAGTGTAGCTAGCAAAAGCTTTATTCAAAGCACTAACCTCATAGTGAAACAAAT	1544A
Qy	1324	TTTAAAGAAATGAGATCATTTGCAAGATTAAGCAATTTTAAACACATCATGTTCCCTGGC	1383G
Dd	1525	TTCAAAGSCAATGTTGTATTTGCAACCATGAGCAATTCGAATGACTTCTTCTCTCGGC	1584A
Qy	1384	ACGGAACCTATCTCTTGTGCTGTGTGTGATGATCTTTTGGTGCACGGACCCCTCTGT	1443G
Dd	1585	ACGGAAGCATATCCCTCGCGCCGCTGGAGATGATTTTGGGCTGTGACCCCTCTGT	1644A
Qy	1444	GATCCAAACGATACGTTTGGCTCCAAAGATGTACATGTTTCATTTTGGCCAAAGCAGC	1503G
Dd	1645	GATCCAAATGGTATATTTTGGCTCAAAGTTGTACATGTTGCTATTTGGCTTACAAAGT	1704A
Qy	1504	TTTGGATGGGAACCTCATCCACCAGATGGGATATGTTCCAAATCCACCACCTCTGT	1563G
Dd	1705	TTATGATGGGAATTTTCATTCATCCAGATGSGGACATGTTTCAGTCCACATCTCTTGT	1764A
Qy	1564	GCCGCTTCCATGCTGCTGCTGAGAGCCATCTGTGTGGCCGAGTCAATGTTAGTATCT	1623G
Dd	1765	GCCGAATTTTCACGCGCGCTCAAGAGCCATCTGTGTGAGCAGTTTACGTTAGTATTC	1824A
Qy	1624	GTTGGAAAGCATTAACCTTGATCTTCTGSAAAAACTAGTCTTCTCGATGAGATCCTT	1683G
Dd	1825	GTTGTATATCAATTTTCAAGTTGTCTCAAAATCTTTTGTTCGCCGACGTTCTATCTTG	1884A
Qy	1684	CGAAGTGTACTATGTGACACGCCGACCTGGGATGTTGTTTGAAGAGCCCTTGGATAT	1743G
Dd	1885	CGTTGTCAACATTAAGCACTGCCCTACACGAGACTGCTTATTTTGAAGAGCCCTTGGATAT	1944A
Qy	1744	GGAGAACTATGCTTAAGATTTGGAATCTTCACACAGTTCACTGAGAGTGAATGGTCAATTC	1803G
Dd	1945	GGCAAAACAATGTCAAAATTTTGGAAATCTCAACAATAATGTGTGTGTTTGGGCTCATTC	2004A
Qy	1804	AACGTCCAGAGAGAGATGCTGTCTGTGAGACACGCCGCAACCACTGTTTTCACAATAC	1866G
Dd	2005	AACGTCCAAAGTGTGTGGTGTGTCTGTGACACGCCGCAAAAGAGCCCTCTCAATATTT	2066A
Qy	1864	TCAAAAGAGTGAATCCAAACCTAACCCAAAAGCATATGATGCAACAGTGGAGAAAC	1922G
Dd	2065	TCAACGCGGAGCATTTTATGCAAGTCTGAAGATATTAAGATGGTGCATGTGA AAAAC	2124A
Qy	1924	CTATCTCTATTGAGAGCGTTTAAACCTTTGGCTTTACCCTGATATACAGCCAAAAAATT	1983G
Dd	2125	CCAATGAGACATCAAAAGTGTGTGATTTTTCGCTGTATATTTTTCAGAGAAAGAACTG	2184A
Qy	1984	ATCTCTTCCAAGCCCTCTCAAGATCTTGACATAGTCTTTGACCACTTGCATTTGCACTC	2043G
Dd	2185	AGTCTCATAGAGTGTTCGATAGATTGGAGATTTTCCCTCGAGCACTTATAGTTTACGCTA	2244A

QY 1390 GCTATCTCTTCGTCGTGTGGTGATGACTTTTGGTGACGAGCCCTCTGGTGATCCA 1449
 1547 GCATATCCCTCGCGCGCGTGGAGATGATTTTGGTGGTGTGATCATCTGTGTATCCA 1606
 QY 1450 AACGTAACGTTTGGTCCCAAGATTCACATGCTTTCATTTGCCCAACGACGCTTGTG 1509
 1607 AATGTAACATATTTGGCTCCAAAGGTTTGTACATGTGATTTGCTTCAACCATTTATGTG 1666
 QY 1510 ATGGGAACATCTACACCTGTGCTGATGATGCTTCAATCCACCCCTTGTGGCGGC 1569
 1667 ATGGGAATTTCTATAGCCAGATTTGGAGATTTTCAAGTCCATCTGCTGTGAA 1726
 QY 1570 TTCCATCTGCTCTCTGAGGACATCTGTGTGGCCGATCTATGTAGTATTTCTGGGA 1629
 1727 TTTCATGCGCGCTCAGAGGACATATCCGGGACCAATTTATTTAGTATTTGTGTGGT 1786
 QY 1630 AACGTAACATTTGATCTTCTGAAAAAAGTATGCTTCTGATGATGATGATCTTCAAGT 1689
 1787 AATGTAACATTTCAAGTGTGCTCAAAATCTTGTGTTTGGCCGATGCTTCACTTGTGCT 1846
 QY 1690 GATGATGATGATCTCCGAGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1749
 1847 CACATTTAGGCACTCCCTACAGAGATTTGCTTGTGAGAGACCTTGTGATTTGCAAA 1906
 QY 1750 ACTATGCTTAAATTTGGAATCTCAACAGTTCATGATGATGATGATGATGATGATGATGAT 1809
 1907 ACATGCTGAAATTTGGAATCTCAACAGTTCATGATGATGATGATGATGATGATGATGAT 1966
 QY 1810 CAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1869
 1967 CAAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2026
 QY 1870 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1929
 2027 GCGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2086
 QY 1930 TCTATGTAAGGCTTAAACCTTTTACCTTATCTATCAAGCCCAAAATTTATCTTC 1989
 2087 AGCAACAAAGGCTTGAATTTTGGTGTATTTTTCAGAGAAAGAAATTTGAGGCTC 2146
 QY 1990 TCCAGAGCCCTCAAGATCTTGAATGATGATGATGATGATGATGATGATGATGATGAT 2049
 2147 ATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2206
 QY 2050 GTTTCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2109
 2207 GTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2266
 QY 2110 AACATGCTTAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2169
 2267 AACATGCTTAAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2226
 QY 2170 GTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2229
 2327 GTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2386
 QY 2230 GCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2289
 2387 TGTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2443
 QY 2290 GTTCAAGTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2322
 2444 GTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2476
 DB 2444 GTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2476
 RESULT 11
 AY081645 2395 bp mRNA linear PLN 14-APR-2002
 LOCUS Arabidopsis thaliana raffinose synthase protein (MPO12.13) mRNA,
 DEFINITION complete cds.
 ACCESSION AY081645
 VERSION AY081645.1 GI:20148632

KEYWORDS
 SOURCE Arabidopsis thaliana.
 ORGANISM Arabidopsis thaliana.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 REFERENCE
 1 (bases 1 to 2395)
 Southwick, A., Karl, J., Neumann, G., Nguyen, M., Lam, B., Miranda, M.,
 Palm, C., J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H.,
 Cheuk, R., Chung, M., K., Hayashizaki, Y., Ishida, J., Kamiya, A.,
 Kawai, J., Kim, C., Lin, J., Liu, S. X., Narusaka, M., Pham, P. K.,
 Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinozaki, K.,
 Shinozaki, K., Ecker, J., Theologis, A. and Davis, R. W.
 Direct Submission
 Submitted (27-FEB-2002) DNA Sequencing and Technology Center,
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,
 USA
 COMMENT
 comment 'e-mail for correspondence: arabseq@stanford.edu
 RIKEN Genomic Sciences Center (GSC) members carried out the
 collection and clustering of RAFL cDNAs (RAFL cDNA : 'RIKEN
 Arabidopsis Full-Length cDNA') : Seki, M., Narusaka, M., Ishida, J.,
 Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
 Hayashizaki, Y. and Shinozaki, K.
 The Sak, Stanford, PGEC (SSP) Consortium members carried out the
 sequencing and annotation of the RAFL cDNAs: Nguyen, M.,
 Southwick, A., Karl, J., Neumann, G., Lam, B., Miranda, M., Palm, C., J.,
 Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M., K.,
 Kim, C., Lin, J., Liu, S. X., Pham, P. K., Sakano, H., Shinozaki, K.,
 Yamada, K., Ecker, J., Theologis, A. and Davis, R. W.
 Nguyen, M., (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed
 equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R. W.
 (SSP/Stanford) contributed equally to this work as PIs.
 FEATURES
 source
 1.2395
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /chromosome="5"
 /clone="U17209"
 /note="This clone is in pENTR/SD-topo. This is a cloned
 PCR product using RIKEN clone RAFL06-10-p15 (AY062781) as
 a template
 ecotype: Columbia"
 1.2395
 /gene="MPO12.13"
 1.2352
 /gene="MPO12.13"
 /codon_start=1
 /product="raffinose synthase protein"
 /protein_id="AA10207.1"
 /db_xref="GI:20148632"
 /translation="MASPCLTKSDSGINGVDTEFERLEDSTLLANGOVYLVDPVNV
 TLTSSPYLNDKDVPLDVSAGSIFNLDEPKSHHVASIGLTKNIKPMISIRFVVM
 TTHWVSGNRDIENETQIILIDSGSDSGSGSRPVYLLLEGSRFSFQGED
 DVAIVCEGSESTGSEVTPROIVYVHAGDPPKLVKDAKVIKLVHVMFKLEKSP
 GIVDFGCTMDAPFLYLVNPDGVHNGCYKIDGCGPGLVITDDQSGISGSDISPP
 EGMNTTVAEGEOPRCLIKFEENHKRDYVSPDQNDVCKAVYRDLKDFSTVDYIV
 WHALCGWGLRPEAPLPSTIRPEISPGKLTMEDLAIVK11ETGIGFSPDLAK
 ERYEGHSHLONAGIDGVKVDYIHLIEMLCQKQYGGVGLAKAVFALTSVVKHNGN
 GVIAEMENDEMFLEGTETASISLRVGDWPCDPSGDPYKGLDGLCHVHCAVNSLM
 MGNFLOPDMDFOSTHPCAEFHAAISGCPITISDCYKIDPDLKRLVPLNSIL
 RCERYALPRDLRDLKIDKTYMLKIWLKNTYVIGAFNCGGSGKCEMFRNOCFS
 ECVNTLRATTSKRDVBNMSSPISITANVEERFALPLDSKRLILSLGNDDELTLPEP
 KRELITVSPVITIGNSVRFAPIGLVNMTSGLAIRSLVINDSEVVEGVFAGEFRVY
 ASKPVSCILDEIVDFEYGEDSMVVOVPMWGPDLSSIOYLF"
 BASE COUNT
 ORIGIN 566 a 532 c 602 g 695 t
 Query Match 40.4%; Score 950.6; DB 8; Length 2395;
 Best Local Similarity 67.4%; Pred. No. 1.7e-272;
 Matches 1422; Conservative 0; Mismatches 644; Indels 45; Gaps 4;

[illegible]

OY	1276	TTGGCAAAAGCATATTTCAAAGCAATGACCAAAATCAATAATAATTAACATTTTAAACGAAT	1335
Db	1291	TTTGGCAAGACTTACTTCAAGAGCGTTAACTCCTCAGTAATAAAGCATTTTAAACGGTAA	1350
OY	1336	GGAGTCATTGGCAACTATGAGAAATTTGAACGACATTCATCTTCTGGCACGCAAGCACTATC	1395
Db	1351	GAGCGTTATCGGCACCATGAGCAATGGCAATGACTTTATGTCTTGCGAAGTGAACCATTC	1410
OY	1386	TCTCTTGGTCTCTTGTGTATGACTTTTGGTGCAGGAGACCCCTCTGGTATCAACAGGT	1455
Db	1411	TCTCTTGGTCTCTGGTGTATGACTTTTGGTGCAGGAGTTCATCAGGCGATCCAAACAGT	1470
OY	1456	ACGTTTGGCTCCAGGATATGCATATGGTTTCATTTGGCCAAAGACAGCTTGGTGGATGGG	1515
Db	1471	ACGTTTGGCTACAGAGATGTGCATATGTCTACTGTGTCTACAAACAGTCTCTGGATGGGA	1530
OY	1516	AACTTCATCACCCCTGCATGGGATATGTTCCAATCCACCACCCTTGTGCCGCTTCAT	1575
Db	1531	AAATTCATTCAGCCTGATTTGGGACATGTTTCAATGCACACATCTCTTGCTGGATGGTCCAT	1590
OY	1576	GCTGCTCTCGAGGCATCTCGTGGTGGCCGCACTCATGTATGATGATCTCTGGGAAAGAT	1635
Db	1591	GCTGCTCTCGAGGCATTTCCGGTGGCCGCAATTCATCATGTATGATGATGTGTGGCAAGAT	1650
OY	1636	AACTTGTATCTTTCGAAAAAAACTAGTCTCTCGATGATGATCATCTCTGGAAGTAGTAC	1695
Db	1651	GATTTGTATCTCTTGAAAGCGCTTGTGTTTGGCCAAAGGTTGATTTGGAGTGTGATGAC	1710
OY	1696	TATGCACTCCCGACTCGCCATTTGTGTTTGAAGACCCCTTGATATATGAGAAACTATG	1755
Db	1711	TATGCTCTCCCAACTCGTGACCGTCTCTTGAAGATCCCTTCATGATGAGGAAAAACATG	1770
OY	1756	CTTAAGATTTGGAAATTCACAAAGTTCACGTGAGTATGGTGGCATTCACATGCGCAAGA	1815
Db	1771	CTCAAGATTTGGAACTTGACAAAGTACACTGAGGATTTATGGAGCACTTCACTGTCTAAGA	1830
OY	1816	GGAGATGTGTGTGTGAGACAGCGCCCAACCAATGCTTTTCAATTACTCAAAAAGAGTG	1875
Db	1831	GGAGATGTGTGTGAGACAAACACAGCACTTAACATGTCTTTCGAATGGGTCAACACGTJA	1890
OY	1876	ACATGCAAAACTTAACCAAAAGACATAGAAATGGGACAGTGGAGAAACCCATGCTCATTT	1935
Db	1891	ACCGCACACACAGCCCTTAAGACGTTGAATGGAACAGTGGAGAGCAGCCCATCTCCATT	1950
OY	1936	GAAAGCGTTAAACCTTGGCGCTTACCTCTATCAAGCCCAAAAACTTATCCTCTCAAG	1995
Db	1951	GCAACGCTGGAAGTATGTGTTGTCTTGTCTCAATCCAAAGAACCTTTTGTGTCTGGA	2010
OY	1996	CCCTCTCAAGATCTTGACATAGCTCTTGACCCATTCGAAATTGACACTCATCAGCTTTCA	2055
Db	2011	CTAAGCATGATCTTGAGCTGACTTTTGGAGCCCTTTCAAATTTGACACTATCAGTCTCT	2070
OY	2056	CCAGTGAACCAACTCAATCCAACTTCTACACTTTTGCCCAATTTGGGCTGGTAAACATG	2115
Db	2071	CCGTGTGACCAATTAAGGGTATTAAGTCAAGTCCGGTTTGCCTGATTTGAGACTGGTAAACATG	2130
OY	2116	CTTAACACTATGTGAGCACTCAATCTGTGAGCATATGACGATGACCTAAGCTCACTGAGAG	2175
Db	2131	CTTAACCAAGCGGTGTGATCCCGCTTTGGTGTATTAAGATGA-----ATCTGTTGAG	2184
OY	2176	ATTTGCTCAAAAGGCTGTGTGATGAGCAAGTATTTGCATGCAAAAACCAAGGCGCTTGT	2235
Db	2185	GTCGGAGTTTTTGGTGGCCGAGAGTTTAGGGTTTATGCAATGCAAAAACCTGTGAGCTGC	2244
OY	2236	CGTATTTGATGGGAGGATGTGGGCTTCAAGTATGATCAAGGACCAATGGTGGTGTTCAA	2295
Db	2245	TTAATTGATGGTGAAGTGTGTGATTTGGTTATG--AAGACTCAATGTGATGTGCAAA	2301
OY	2296	GTCGCATGGCC	2306
Db	2302	GTCGCTTGGTC	2312

RESULT 12
AY062781 2718 bp mRNA linear PLN 25-NOV-2001
LOCUS Arabidopsis thaliana raffinose synthase protein (WP012.13) mRNA,
DEFINITION complete cds.
ACCESSION AY062781.1 GI:17065409
VERSION FL1.CDNA.
KEYWORDS Arabidopsis thaliana.
SOURCE Arabidopsis thaliana.
ORGANISM Arabidopsis thaliana.
REFERENCE Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons: core eudicots:
Rostidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 2718)
AUTHORS Southwick, A., Karlin-Neumann, G., Nguyen, M., Lam, B., Miranda, M.,
Palm, C.J., Bowser, L., Jones, T., Banh, J., Carninci, P., Chen, H.,
Cheuk, R., Chung, M.K., Hayashizaki, Y., Ishida, J., Kamiya, A.,
Kawai, J., Kim, C., Lin, J., Liu, S.X., Narusaka, M., Pham, P.K.,
Sakano, H., Sakurai, T., Satou, M., Seki, M., Shinn, P., Yamada, K.,
Shinozaki, K., Ecker, J., Theologis, A. and Davis, R.W.
TITLE Submitted (14-NOV-2001) DNA Sequencing and Technology Center,
Journal Stanford University, 855 California Avenue, Palo Alto, CA 94304,
USA
COMMENT e-mail for correspondence: arabesequence.stanford.edu
RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL cDNA: "RIKEN
Arabidopsis Full-length cDNA"): Seki, M., Narusaka, M., Ishida, J.,
Satou, M., Kamiya, A., Sakurai, T., Carninci, P., Kawai, J.,
Hayashizaki, Y. and Shinozaki, K.
The Sak, Stanford, PGEC (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Nguyen, M.,
Southwick, A., Kerlin-Neumann, G., Lam, B., Miranda, M., Palm, C.J.,
Bowser, L., Jones, T., Banh, J., Chen, H., Cheuk, R., Chung, M.K.,
Kim, C., Lin, J., Liu, S.X., Pham, P.K., Sakano, H., Shinn, P.,
Yamada, K., Ecker, J., Theologis, A. and Davis, R.W.
Southwick, A., (SSP/Stanford) and Seki, M. (RIKEN GSC) contributed
equally to this work. Shinozaki, K. (RIKEN GSC) and Davis, R.W.
(SSP/Stanford) contributed equally to this work as PIs.
FEATURES
source Location/Qualifiers
1..2718
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/chromosome="v"
/clone="RAFL06-10-P15"
/note="This clone is in bluescript
ecotype: Columbia"
1..2718
/gene="MP012.13"
132..2483
/gene="MP012.13"
/codon_start=1
/product="raffinose synthase protein"
/protein_id="AAL32859.1"
/db_xref="GI:17065410"
translation="MASPLTKRSDSGINGVDFTEKFRLEDSILLANGVLTLDVENV
TLTSPYLVDKGVPLDYASGIFGINDGPKSHVAVSLGKTKIRMSIFRQVVM
TTHWGSNGRIENETQIILIDSGSDSGPSGSRPVLPIRLLEGSFRSSFOGED
DDVAVCESGTEVESEFRQIVYVHADGPKLKDAMKIRIVAMTFKLEESPP
GIVDFGCTWDAPFLYVNPAGVHKVGLVDGCPPLVLDGMSIGHSDSIDV
EGMTVAGDEMPKRLKEENHKFKDYVSPKDDVMKAFVRDLKDFEFVVDIYV
WHALCGVGLRPEAPALPSTIIRPELSPGLKLMEDLAVDKITETGIGASPLAK
EYEGLSHLNAGIDGKRVYIHLLELCKQYGRVLAATFRALTSVKNKRNNG
GVASHEKNDPMELGTALISLGRVDFWCTDPSDGVGKHFDLRLVLPNSIL
MGNFIQDPMDFQSTHPCAEFRAASRAISGPIYISDCVNGKHFDLRLVLPNSIL
RCEYALPTDRLEFDPLHDKTLMKINLNKYTGIVGAFNCGGCMGCRTRRNCFS
ECVNTLATPTSKDEVMSGSPISIAVVEEPAFLFLOSCKLLSGLDLDELTEPF
KPELTIVPVTITGNSVPAFICLVNMLNLSGATRSIYVNDSEVGVFCGGERVY
ASKRVSCLLIGEVFEREDSMVAVQVWSPGDLSSIQYLF"

BASE COUNT

658 a 581 c 643 g 836 t

ORIGIN

Query Match 40.4%; Score 950.6; DB 8; Length 2718;
Best Local Similarity 67.4%; Pred. No. 1.8e-272;
Matches 1422; Conservative 0; Mismatches 644; Indels 45; Gaps 4;

QY 220 GAACCTGATAGCCGACATGTTGTTGATTTGGGAAGCTGAAAGATATTCGTTTATGAGT 279
DB 354 GAGCCCAAAAGCCACACAGTGCATTCATCGAAAACCAACATTCGTTTCATGAGC 413
QY 280 ATTTTCAGGTTTAAGCTTTGGTGACACTGACACTGGGTTGGTGAATATGCGGATCTT 339
DB 414 ATATTCGTTTCAAGTTTGGTGAGTACTCATTTGGTGATCAAAACGAGCTGACATC 473
QY 340 GAATCGAGACATCAGATTGTGATCTCTTGAGAA-----GTCA 375
DB 474 GAGAACGAGACTCAATATCTTCTTGAATCAATCCGGGTCGATTTGGACCCGGATCC 533
QY 376 GATTCGTGCGACCGATGATTTCTTCTTCCGATTCGTTGAGGAGCCGTTCCGAACCTCG 435
DB 534 GGGTCGGGCGTCTTATGTCTCTTGGTACCTCTTGAAGGCGCTTCCGTTTCATCA 593
QY 436 ATTCAGCCTCGGAGATGATGACTTTGTCATGTTTGTGTCGAGATGTTTCGTCGAAGTT 495
DB 594 TTCCATCCGGAAGACATGACGTGGCGGCTGTGTGTAATCCGGGTCGACCGAAGTA 653
QY 496 GTTGATGCATCTGTCGGAAGTATGTTGTAATCTTCATGCTGGTGAATGATCCGTTGCAT 555
DB 654 ACCGGTCGAGTTTCGTCAGATTTGTAATGTCATGCGGGGATGACCCGTTCAAGCTC 713
QY 556 GTTAAAGAGCGCATAGATTCGTGAGACCCATCTTGGAACTTTTCGTTTGGAGAGG 615
DB 714 GTGAAGAGCCCATGAAATGATTAAGGTTTCAATATGATTAAGCTTCAAGCTTCAAGAG 773
QY 616 AAGACTCCACCGATGATGTCGACAAATTCGTTGGTCGACGTGAGAGCGCTTTTACCTA 675
DB 774 AATGCGCCCGCGGAATGTCATTAATTCGGGTGTCACATTTGATCTTTTACCTG 833
QY 676 ACGGTTCATCCACAGGCGGTATAGAAAGCGGAGCATCTGTGAGCGGCTGTCTCT 735
DB 834 ACGGTGAATCTCGACGAGATTCATTAAGGTGTTAAGTCTCTGACGCGTGTGTCTCT 893
QY 736 CCGGTTTATGTCCTATACGATGATGTTGGCAATCCATCGACAGATTCGGATCCATC 795
DB 894 CCGGATGTCCTTATGACGACGATGGAATGATGATTTGACATGATTCGATGATAT 953
QY 796 ACCAAGAAGATGACCAACACGTCGCGCGGACGAACATCCCTGCTTTTGA 855
DB 954 GATGTCGAAGGATGATATTAACCGTCGCGGTCGAACAAATGCTTGACGCTTGAG 1013
QY 856 TTCCAAGAGATTAACAATTCGCTGACATGTCATCCCAAGGCCACGCGCCCGAGCC 915
DB 1014 TTTGAAGAGAACCAATTAACAAGCTTCTTCTCCAAAG-----ATCAA 1061
QY 916 GGCAGAAAGGAGATGAAGCGTTTATAGATGAATCAAAAGAGAGTTTAAAGCTGTGAG 975
DB 1062 AAGCAGCTCGGAATGAACCTTTCGACAGATGCAAGATGAATTCACACCGTTGAT 1121
QY 976 CATGTTTATGTTTGGCATCTTTGTGTGATATTTGGGGTGGCTTCCCGCAGATGCTT 1035
DB 1122 TGCATCTTACGTTTGGCACCCATTTGCGGTTTCTGAGGAGTCTCCCTTAAGCTCGC 1181
QY 1036 GCGTTGCTCGAGGACAGTGTGATTAACGACGAGCTTTCACGAGGCTGAGATGACGATG 1095
DB 1182 GCTTCGCGCCGTCGACCATTTATCCGCGCCGAGCTCTCTCCAGACTTAACTAACATG 1241
QY 1096 GAGATTTGGCGGTGATTAAGATTTGTTTCAATAGCTCGGCTGTCCCGCGGAGAG 1155
DB 1242 GAAAGATCTCGCGGTGATTAAGATCAATAGAGACCGGAATGGATTTGGCTCGCGGAG 1301
QY 1156 GCTGAGAGATATGACGAGACTTCACTGCTCATTTGGAAAAAGTTGGATTCACGCTGT 1215
DB 1302 GCGAAGAGTTTACGAAAGGTTTCACTCTCATTTCAAAAGCGCGGTATTTGACGCGCTT 1361

Qy	1216	AAAGTTGACGTTATCCCAACCTATTGGAGAGATTGTGTGAAGACTATGAGGAGGAGACTGGAT	1275
Db	1352	AAAGTTGATGTGCATCCACATATTGGAGATGTATTGCGCAAAATATTGGCGGAGACTTGC	1421
Qy	1276	TTTGCAAAAGGCATATTACAAAGCAATGACCMAATCAATAAATMAACATTTTAAAGGAAT	1335
Db	1422	TTTGCCCAAAAGCTTACTTCACAGCGGCTTAACGTCGTGAGTAATMAACATTTTAAACGGTAAAC	1481
Qy	1336	GGACTCAATTCGAAGTATGSAACATTTGAACGATTCATGTCCTTGCCACGGAACCTATC	1395
Db	1482	GGCGTTATCCGACACATGAGGACATTCGAATGATTTATTTCTCTTGGAACGTGAACCCATC	1541
Qy	1396	TCTCTTGGTGGTGTGGTGTGATCTTTTGGTGCACGAGACCCCTCGATGATCCAAACGCT	1455
Db	1542	TCTCTTGGTGGTGTGGTGTGATCTTTTGGTGCACGAGATTCATCAGGGGATCCAAACGCT	1601
Qy	1456	ACGTTTGGCTCCCAAGATGTCACTATGTTTCATTTGGCCACGACGACTTGGATGGGG	1515
Db	1602	ACGTTTGGCTCCCAAGATGTCACTATGTTTCATTTGGCCACGACGACTTGGATGGGG	1661
Qy	1516	AACTTCATCCACCCCTGCATGGATATGTTCCATATCACCACCCTTGTGCGCCCTCCAT	1575
Db	1662	AACTTCATCCACCCCTGCATGGATATGTTCCATATCACCACCCTTGTGCGCCCTCCAT	1721
Qy	1576	GCTGCTCTGTGAGCCATCTGTGTGGCCGATCTATGTTAGTATTCGTGGGAAAGCAT	1635
Db	1722	GCTGCTCTGTGAGCCATCTGTGTGGCCGATCTATGTTAGTATTCGTGGGAAAGCAT	1781
Qy	1636	AACCTTATCTCTGAAAAAAACATAGAGCTTCCGATGAGTGCATCCTTCGSAAGTAGATAC	1695
Db	1782	GATTTTATCTCTTGAAGGCTCTGTGTTGGCCCAACGGTTCATTTTGAAGTGTGATGAC	1841
Qy	1696	TATGCACTCCCGACTCGCATGTTTGTGTTGAAGACCCCTTGCATATGAGAAACATAG	1755
Db	1842	TATGCTCTCCCAACTCGTGTGACCCGCTCTTGTGAGAGATCCCTTCATGATGAGCAAAACCATG	1901
Qy	1756	CTTAAGATTGGGATCTCAACAGTTCACTGAGAGTATGGTGCATTCACATGCCAAGA	1815
Db	1902	CTCAAGATTGGGATCTCAACAGTTCACTGAGAGTATGGTGCATTCACATGCCAAGA	1961
Qy	1816	GGAGAGTGTGTGTGTGAGACACCCCAACCAATGCTTTTCAACATACTCAAAACGAGTG	1875
Db	1962	GGAGAGTGTGTGTGTGAGACACCCCAACCAATGCTTTTCTGATAGTGGTCAACACGTTA	2021
Qy	1876	ACATCCAAAACCTAACCCAAAACATAGATGAGACAGTGGAGAAACCCATATCTTATTT	1935
Db	2022	ACCCCAACCAACAGCCCTAAAGACGTTGATGGAACAGTGGAGACGCCCAATCTCCATTT	2081
Qy	1936	GAAAGCCTTAAACCTTTGGGCGTTTACGCTCTTTCACAGCCCAAAAACCTATCCTCCAAAG	1995
Db	2082	GCAACGTTTGAAGAGTTTGGCTTTGTTCTTGTCTCAATCCAGAACCTTTTGTGTGGA	2141
Qy	1996	CCCTCTCAAGATCTTGACATAGCTCTTGAACCCATTCGAATTCGAATTCACATGTTTCA	2055
Db	2142	CTAAGCATGATCTTGAGCTGACTTTTGGAGCCCTTCAAGTTTGAAGCTATCATGCTGCT	2201
Qy	2056	CCAGTACCAAACTCATCAAACCTTCTTCAACCTTTGCCCAATTTGGCTGTGAACATG	2115
Db	2202	CCGTTTGTGACCATTAAGGGTAAATTAAGTCGCGGTGTGCTCCCATTTGACTGTTTAAACATG	2261
Qy	2116	CTTAACACTATGTGAGCATCCCAATCTGTGAGCATGTGAGATGACCTTAAGCTCAGTGGAG	2175
Db	2262	CTTAACCAACAGCGGTGCGATCGCGTCTTTGGTGTGTATATGATGA-----ATCTGTTGAG	2315
Qy	2176	ATTGGTGTCAAAGGCTGTGTGTGATGCGAGTATTTGCATGCAAAAACCAAGGCTTGT	2235
Db	2316	GTCGAGATTTTGTGTGCGGAGAGATTAGGGTTTATGATCATGAAAACCTGTGAGCTGC	2375
Qy	2236	CGATTGATGGGAGAGATGTTGGCTTCAAGTATGATCAGAGCAAAATGCTGTGCTTCAA	2295
Db	2376	TTAATTTATGTGAAGTGTGTGAGTTTGGTTATG---AAGACTCAATGATGATGCTGCA	2432

[illegible]

D	361	GC	CC	CC	AGGGGTTATG	AG	GGGCGCTG	CA	GGGGCGT	GT	GT	GA	CGGCGGAT	GT	CGCGGGCGCT	420																			
O	744	AG	TC	CT	TA	TG	CA	GC	AG	TG	TTG	GC	CA	T	CG	CA	AT	CG	CA	CG	AT	T	CC	CA	T	CA	CA	AA	GA	803					
D	421	GG	TT	GT	AT	G	CA	GC	AG	CG	GGG	GGG	CG	AG	T	CC	AT	TT	T	T	T	T	T	T	T	T	T	T	T	T	480				
O	804	AG	CA	AT	GA	AC	CA	AA	CC	CT	CG	CC	CG	CA	GC	CA	AT	TG	CC	T	CG	CT	CT	TT	T	GA	AT	T	CA	GA	863				
D	481	GG	GG	AT	GGG	GA	AA	CC	T	CG	CC	CG	GA	AG	CA	AA	T	GG	CC	T	CG	CA	AG	TT	GA	T	CA	AA	T	GA	GA	540			
O	864	GA	AT	T	CA	AA	AT	T	CG	TA	C	AG	T	CA	AT	CC	AA	GG	CG	CC	CG	CC	CG	CG	CG	CG	CG	CG	CG	CA	GA	923			
D	541	GA	AT	T	CA	AA	AG	T	CA	GC	AG	AG	T	CA	GC	AG	CC	CG	AA	T	TA	AA	CT	GG	GG	CC	---	GG	CC	CG	AA	T	AC	597	
O	924	GG	GG	AT	GA	AG	CC	GT	TT	A	T	GA	T	GA	CT	CA	AA	GG	AG	AG	T	T	T	AA	AG	T	GA	AG	T	GA	AG	CA	T	GA	983
D	598	GG	GG	AT	GG	GG	CC	CT	TT	A	T	T	CG	T	GA	C	A	T	GA	AA	GG	C	A	T	TT	CA	AG	AG	T	GA	AG	CT	GA	957	
O	984	TG	TT	T	GC	A	T	GC	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1043		
D	658	CG	T	GT	GC	A	T	GC	A	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	717		
O	1044	TG	AG	CA	CG	TG	TA	T	GA	GC	CA	T	GC	TT	T	CA	C	AG	GG	CG	T	GA	C	AG	AT	GA	CG	AT	GA	AG	AT	T	1103		
D	718	CG	AG	GC	CT	TA	GC	CT	A	TT	GA	CC	CA	AA	CT	GA	CT	CT	TG	GG	CT	TT	GA	AG	CA	CC	AT	CA	T	GA	AG	AT	777		
O	1104	GG	CG	T	GA	AT	PA	AG	T	T	GT	TT	CT	T	CA	T	AA	GG	T	CG	GG	CT	GC	T	CG	CC	CG	GA	AG	GG	CT	GA	GA	1163	
D	778	GG	CT	GT	T	GA	T	GA	AT	GT	T	CA	AA	CA	AT	T	GG	CG	GG	GT	CT	GT	GT	CC	CA	CG	GA	AT	TT	GT	T	GA	CA	837	
O	1164	GAT	GT	AC	GA	AG	AG	CT	T	CA	T	GC	AT	TT	T	GA	AA	AG	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1223		
D	838	AAT	GT	AT	GA	AG	AG	AT	T	CA	AT	T	CA	AT	CT	CG	A	T	CT	T	GG	GG	AT	T	GA	T	GA	T	GA	T	GA	T	GA	897	
O	1224	CG	TA	T	GC	AC	CA	CT	AT	T	GA	AG	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1283		
D	898	CG	T	AT	CA	CT	AT	TT	GT	T	GA	AA	T	GT	T	GT	GA	AG	CT	A	T	GG	T	GG	AG	AG	AG	AG	AG	AG	AG	AG	957		
O	1284	GG	CA	T	AT	T	CA	AA	GC	AT	T	GA	C	AA	T	CA	A	T	AA	T	AA	CA	CT	TT	T	T	T	T	T	T	T	T	1343		
D	958	GG	CT	AT	AT	T	CA	AA	GG	CC	TT	AT	T	CA	AA	GC	T	AG	T	T	TA	CA	CA	CT	TT	CA	AG	GA	AG	GG	CT	AT	1017		
O	1344	TG	CA	AT	T	GA	AG	AG	CT	T	T	GA	CA	CT	T	GT	CC	T	T	GC	AC	CG	GA	AG	CT	AT	T	CT	T	CT	T	GG	1403		
D	1018	CG	CG	GC	CT	GG	AG	CA	CG	CA	AT	CA	CT	CA	T	GT	TT	T	CG	GA	CC	GA	GC	CA	CT	AT	CA	CT	T	GG	1077				
O	1404	TG	CG	T	GT	TG	TG	AT	C	AT	TT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1463			
D	1078	TG	CT	GT	CG	GG	AT	AT	TT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1137			
O	1464	GCT	CA	AG	AT	GC	AT	C	A	T	T	GC	TT	AT	T	T	GC	CA	CA	GC	A	CT	T	T	GG	AT	GG	GA	AG	AT	TC	AT	1523		
D	1138	GTT	CA	AA	GG	GT	CA	AT	T	GC	AT	T	GC	AC	T	GG	CG	CT	CA	CA	CA	CA	AT	T	AT	T	AT	T	AT	T	AT	1197			
O	1524	CC	AC	CC	T	GA	CT	GG	GA	T	AT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1583			
D	1198	CC	AC	CC	T	GA	AT	T	GG	CA	AT	GT	TT	CA	T	CA	T	CA	CT	C	AC	CT	T	CG	CT	GA	AT	T	CC	AC	GC	CT	1257		
O	1584	TG	CA	GC	CA	T	CT	CG	TG	GG	CC	CG	AT	CT	AT	T	GA	T	AT	T	CT	T	GA	GG	AA	AG	CA	TA	CT	T	GA	1643			
D	1258	AC	GA	GC	CA	T	CT	CG	GG	CG	CC	CA	TT	T	AC	GT	CA	CT	CG	T	GG	T	CG	GA	AA	CA	CA	AA	CT	T	GA	1317			
O	1644	TG	TT	T	CA	AAAA	AT	T	GA	TG	CT	T	GC	T	CA	T	GA	T	GA	T	GA	T	GA	T	GA	T	GA	T	GA	T	GA	1703			
D	1318	GCT	CT	T	TA	AG	AG	CT	CT	TT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1377			
O	1704	CC	GC	AT	CG	CG	AT	T	GT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1763			
D	1378	TG	CA	CT	AC	TG	CA	T	T	GC	CT	TT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1437			
O	1764	TTG	GA	AT	CT	CA	AA	AG	T	T	CA	CT	GA	GT	AT	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	T	1823			
D	1438	TTG	GA	AT	T	TA	CA	AA	AG	T	T	CA	CG	AG	T	T	CG	AA	CT	T	T	CA	CT	T	GC	CA	AG	GT	T	GC	GG	GT	1497		

QY	1824	GGTGTGTGGACAGCCCGCAACCATGCTTTTACAAATCTCCAAAAGAGTGACATTC	1883
Db	1498	GAGCGGGGAGTGTGCTCGCAACCAATTCGCTCGGAGTATTTCCACGCCGTTCTCTAG	1557
QY	1884	AACATACCCAAAAGACATAGATGGCACAGTGGAGAGAAAACCTATCTCTATTGAAGCGCT	1943
Db	1558	CGCTGTGCGAGTGAATGAGTGGAGACAAAGACGATCGATCGAGTGCAGCGGCT	1617
QY	1944	TAAACCTTTGGGCTTTACCTCTATCCAGCCAAAACCTATCTCTCTCAACCCCTCTCA	2003
Db	1618	CAAAACATTGCGCTGTACCTATTCACACAGAGAAACCTGCTCTTTTAACCCATTCACA	1677
QY	2004	AGATCTTGACATAGCTCTTTGACCCATTGCAATTCGACGTCATCACTGTTTCCACAGTAC	2063
Db	1678	CAAAATTCGACATCAGCGCTTGAGCCCTTGATTTTGAGTGATTAACCGTTTCTCCAGTCA	1737
QY	2064	CAAACT 2069	
Db	1738	AACTCT 1743	

RESULT 14				
E36418	E36418	1762 bp	DNA	linear
LOCUS	Raffinose synthase gene.			
DEFINITION	E36418			
ACCESSION	E36418.1			
VERSION	GI:13019216			
KEYWORDS	JP 2000014389-A/2.			
SOURCE	Brassica napus.			
ORGANISM	Brassica napus			
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae; Brassica.			
REFERENCE	1 (bases 1 to 1762)			
AUTHORS	Eljaro, W. and Kenji, O.			
TITLE	Raffinose synthase gene			
JOURNAL	Patent: JP 2000014389-A 2 18-JAN-2000;			

COMMENT
OS BRASSICA napus
PN JP 2000014389-A/2
PD 18-JAN-2000
PF 10-DEC-1998 JP 1998351246
PR
PI EIJIRO WATANABE, KENJI OEDA
PC C12N15/09, A01H5/00, C12N1/21, C12N5/10, C12N9/88, C12Q1/68// PC
(C12N1/21, C12N1:19), (C12N5/10, C12R1:91), (C12N9/88, C12R1:19), PC
(C12N9/88, C12R1:91), C12N15/00, C12N5/00, (C12N5/00, C12R1:91) CC
FH Key location/Qualifiers
FT CDS (1). (1719).

FEATURES	LOCATION/Qualifiers
SOURCE	1. :1762 /organism="Brassica napus" /db_xref="taxon:3708"
BASE COUNT	435 a 384 c 459 g 479 t 5 others
ORIGIN	
Query Match	36.8% Score 866.4; DB 6; Length 1762;
Best Local Similarity	70.6% ; Pred. No.2.6e-247;
Matches 1200: Conservative	4; Mismatches 475; Indels 21; Gaps 3

QY	607	TTTGAGGAGAAAGACTCCACAGATATCGTGACAAATTTGGTTGGTGCACGTGGAGCCG	666
Db	1	TTGGAAGAAAAAACGCCGCCGGGAATCGTCATTAAGTTCCGGGTGTGCACGTGGATGCG	60
QY	667	TTTTTACCTAAGCGTTTCATCCACAGGCGCTAATAGAAGCGTGAAGCATCTTCGTGACGCG	726
Db	61	TTTTTATTTTGACGCGTAAACCCCTGACGCGAGTTTCATAAGGCTGTTAAGTCTCTGTCGACGGT	120
QY	727	GGTTTCTCCCGGTTTAACTCTAATGCACATGCGTTGGCAATCCATCGGACACAGATTGC	786
Db	121	GGTTTCTCCCGCGGATTTGTCCTATGCACAGCGTTGGCAATCGATTGGAATGATCTCC	180

Oy	787	GATCCCATACCAAGAAAGAAATGAACCAAAACCGTCGCCGGGAGCAAAATGGCCCGGCT	846
Db	181	GATGATTCATGATTGTTGAAGGATGATGTTGTAACCGTCGCCGGGAGCAAAATGGCTTGACAG	240
Oy	847	CTTTTGAATTTCCAAAGAAATTTACAAATTTCCGTAGCTACGTCAATCCCAAGGACACCGGC	906
Db	241	CTTCCGAAATTTCAAGAGAACTTCAAGTTCCAGAGACTACGTCTCTCCGAAACAC-----	295
Oy	907	CCCCAGCCGGCCAGAAAGGGGATGAAGGCGTTTATAGATGAACCTCAAGAGAGATTTAAG	966
Db	296	-----AAAACGAAGTCGGGATGAAGAAAGCTTTCTGCAGAGATCTGAAGAAATTTCTCC	348
Oy	967	ACTGTGAGGCAATGTTTATGTTTGGCAATGCTTGTGTGATATTTGGGTGCGCTTCGCCGC	1026
Db	349	ACCTGTATTTACATCTACGTCGTGGACGGCGTTTGCGGTACTGGGGGWTGTTGTGTCC	408
Oy	1027	CAGTGGCTGGCTTGCCTGAGGCACTGTGATTCACCAAGTCTTTACACAGCGCTTCAG	1086
Db	409	GGACCTCTTACTCTTCGCCCTGACTACTTGTCCGACAGACTCTCGCGGGGCTTTAAG	468
Oy	1087	ATGACATGAGAGATTTGGCGGTGATTAAGATTTGTTCTTATTAAGTGGGCTGTCCG	1148
Db	469	TTGACATGACAGATCTCGCCGTTGATTAAGATCATGATACCGGAATCGATTCTGTCCG	528
Oy	1147	CCGAGAGAGCGTGAGAGATGTACGAAAGCACTCATGCTCATTTGGAAAAAGTTGGGATC	1206
Db	529	CCGACATGGGAAACGAGTTTATACGAAAGCTTCACTGTCAATCTTCAAAACGTCCGGCAT	588
Oy	1207	GACGCTTTAAGATTGACGTTATCCACCTATTGGAGATTTGTGTGAAGACTATGAGAG	1266
Db	589	AAGCGCTTAAAGTTGACGTTATCCACATACGTGANAATTTGTGGAGAAATATGGCGG	648
Oy	1267	AGAGTGATTTGGCAAAAGCATTTACAAAGCAATGACCAAAATCAATAATTAACATTTT	1328
Db	649	AGAGTTACTTGGCTAAAGCTTACTTCAAGGCGCTTAACGCTGACGTAAAGCATTTT	708
Oy	1327	AAAGGAATGAGCATTTGAAGTATGAAACATTTGAAGACTTATGTTCTTGGCACG	1386
Db	709	GACGGCAACCGCGTTATTCGCGACATGAGACACTGTAAAGACTTATGTTCTTGGAAAC	768
Oy	1387	GAACTATCTCTTGGTGCCTGTTGGTGAATGATTTGGTGCACGAGCCCTCTGTGAT	1446
Db	769	GAACTATCTCTAGAGTCTGTGCGGTGATGACTTTGGTGCACGATCATCTGGCAC	828
Oy	1447	CAAAGCGTACGTTTGGCTCCAAAGATGTACATGTTCAATTTGTGCCAAGCACAGCTTG	1506
Db	829	ATTAAACGGCAGCTATTGGCTGCAAGAGATGACATGTCCTACATGACCTAACACAGCTT	888
Oy	1507	TGATATGGGGAATTCATCCACCCCTGACTGGGATATGTTCCATTCACCCACCTCTGTGC	1566
Db	889	TGATATGGGAATTTTATCCACGCTGTATGGGAATTTTCACTCCACACATCTTGTGCT	948
Oy	1567	GCTTTCATGCTGCTCTCAGCCATCTCGGGGCGCCATATGTTAGTATCTGTG	1626
Db	949	GAGTTCATGCTGTTTACGTGTCCGATCTCCGGTGGGCCATTTATCATCAGCATTTGTG	1008
Oy	1627	GGAAGCATAACTTGTATCTTCTGAAAAAACTAGTCTTCTGATGATTCGATCTTGA	1686
Db	1009	GGCCAGACATTTTCGATCTTCTTGAGAGACCTGTTTCTGACGGTTCGATTTGAGG	1068
Oy	1687	AGTAGATCATGTGATCCCGGACTCGCGATTTGTTTGTGAAGACCTTTTGATATATGA	1746
Db	1069	TGTGAGTACATGCTCTCCCACTCTGACCGTCTTTTGAAGACCCCTTCATCATATGC	1128
Oy	1747	GAACTATGCTTAAGATTGGAATCTCAACAAGTTCACTGAGATTAAGTGGTCATTAC	1806
Db	1129	AAAACCATGCTCAGATTTTGAATCTTGAACACAGTACACTGTGATCATGTGGACATTAC	1188
Oy	1807	TGCCACGAGAGAGATGTGTCTGTGAGACACCGCCGACCAATGCTTTTCAACAATACCA	1866
Db	1189	TGTCAAGAGAGAGATGTGTCAAGAAACTCAGCGGACCAATGCTTCTCCCAATGCGTT	1248
Oy	1867	AAACGATGATCCAAACTAACCCAAAAAGACATAGATGGCACAGTGGAGAAAACCT	1926

[illegible]

RESULT	15
PSA311087	
LOCUS	PSA311087 2727 bp mRNA Linear PLN 03-JAN-2002
DEFINITION	Pisum sativum mRNA for stachyose synthase (stsl gene).
ACCESSION	AJ311087
VERSION	AJ311087.1 GI:13992584
KEYWORDS	stachyose synthase; stsl+ gene.
SOURCE	pea.
ORGANISM	Pisum sativum
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
REFERENCE	1
AUTHORS	Peterbauer,T., Mucha,J., Mach,L. and Richter,A.
TITLE	Chain Elongation of Raffinose In Pea Seeds. ISOLATION, CHARACTERIZATION, AND MOLECULAR CLONING OF A MULTIFUNCTIONAL ENZYME CATALYZING THE SYNTHESIS OF STACHYOSE AND VERBASCOSE
JOURNAL	J. Biol. Chem. 277 (1), 194-200 (2002)
PUBMED	11675396
REFERENCE	2 (bases 1 to 2727) Peterbauer,T. Direct Submission
AUTHORS	Submitted (02-MAY-2001) Peterbauer T., Institute of Ecology, University of Vienna, Althanstrasse 14, A-1090 Vienna, AUSTRIA
JOURNAL	Location/Qualifiers
FEATURES	1..2727
SOURCE	/organism="Pisum sativum" /cultivar="Wunder von Kelvedon" /db_xref="taxon:3888" /tissue_type="seeds" /dev_stage="developing seeds" /country="Austria" 57..2618 /gene="stsl" 57..2618 /gene="stsl" /EC_number="2.4.1.67" /codon_start=1 /evidence=experimental /product="stachyose synthase"
gene	
CDS	

```

/protein_id="CAC38094.1"
/db_xref="GI:13992585"
/translation="MAPLNSTSTNLKTESIFDLSEKFKVKGPFPLPHVDENVSR
SSSICKPSESNAPSLLOKYLAYSXKSGFSGFSEMDLSEKSNISGFNKDELISIF
RPFKMSQWIKGSGSDQMETONLILHVPETKSVVLIIPLEKFRALPPGPDHV
KIMBSGTAKVESTFNIAVHSENYDLMEKYSVIRHLSFRALPERKTLPLV
DLFWCWDADAVLYLVNPIGIFHGLDSEKSGVEPFIIDGWMOSISPDGJHEKLR
NLVLGEGMSGRLHRFDECFKFRKESGLLGPNSPYDPNNFTDLILKJIEHEKLR
KRELAISKSSDLAEIESKIKKVVIEDLFGSGEBSSEKSEKSEGLAKFTKDLR
TKFKGLDVLVYHMLCGAMGVREPETHLDKRYDCKLSPGLDGMEDLAVVEYSKAS
LGIWPSQANLXDMSHSYLAESCITGVYKIVHLSVECDYGRDLAVVEYEGLT
KSIYKFNNGMISMQHCNDPFFELGTRQISMGRDIPFNDPDEGDSFMILOGVH
MHCYSNLSLMGOMIQPDMDFSDHVCARHAFSRAICGPIYSDVNGSHDEDLK
KLVPDGTIPKCIYFPLPTRDCLFKNPLFDHTTVLKIMFNKRGVIGAFNCGAGMD
PIOMKRGFPECYKPICTVHTEVEMQKEETSHLGAEEVYVLYNLAELSMLTK
SEPIOTFOTSPFELSPVPTKLOGGIKFAPIDGTYNNNSGQVYIDLEYGNGLKIK
VKGGSFIAISSESPKRLQNLGCEYDFEMLDGLCLVAVPIIEACVSDMEIEF"

BASE COUNT      809 a      417 c      584 g      917 t
ORIGIN
Query Match      17.3% Score 406.2; DB 8; Length 2727;
Best Local Similarity 57.7%; Pred. No. 1.5e-109;
Matches 808; Conservative 0; Mismatches 578; Indels 15; Gaps 4:

QY  920 AGAAGGGATGAAGCGCTTATAGATGAACCTCAAGAGAGATTAAAGCTGGAGCAGT 979
    || || || || || || || || || || || || || || || || || || || ||
DB  1198 AGTATGATTAAGAGGCTTACACAAAGATTGAGGACAAAGTTCAAAAGTTGGATGATG 1257

QY  980 TTATATGTTGGCATCTTGTGTGATATTGGGCGTGGCTTGGCCGCGAGTGCCGTGCT 1039
    || || || || || || || || || || || || || || || || || || || ||
DB  1238 TTTATATGTTGGCATCTTGTGTGATATTGGGCGTGGCTTGGCCGCGAGTGCCGTGCT 1317

QY  1040 TGCCTGAGGACGTGATTCAGCAGTCTTTCACACGAGGTGAGATGATGAGGAGT 1099
    || || || || || || || || || || || || || || || || || || || ||
DB  1318 TGTATACCAAAATGTGCCCTTGCAAA---CTCTCACTGCTGTGATGAGACATGAGAG 1374

QY  1100 ATTTGGCGGTGATAGATTGTTCTTCATAGGTGCGGCTGCTCCCGCGAGAGAGCTG 1159
    || || || || || || || || || || || || || || || || || || || ||
DB  1375 ATCTTGCACTGGTGTGATTTCCAAAGCGTCACTTGGGTAGTTCCTTCTAGTCAAGCTA 1434

QY  1160 AGCAGATGTACGAAGAGACTTCATGCTCATTTGGAAAAAGTTGGAGATCGAGGGTAAAGA 1219
    || || || || || || || || || || || || || || || || || || || ||
DB  1435 ATGAGCTTTATATGATTCATCTTCTTATCTTGCTGAAATCCCGTATCACCGGAGTAAAG 1494

QY  1220 TTGAGCTTATCCACCTATTTGAGATGTTGTGTGAACACTATGAGAGGAGTGGATTGG 1279
    || || || || || || || || || || || || || || || || || || || ||
DB  1495 TTGATGTCATTCATCTCTCTGTAATATGCTGTGATGAAATGAGAGTAAAGTTGATCTTG 1554

QY  1280 CAAAGCATATTACAAAGCAATGACCAATCAATTAATTAACATTTTAAAGCAATGAG 1339
    || || || || || || || || || || || || || || || || || || || ||
DB  1555 CAAAAGTTTACTATGACAAAGATTGACAAATTCATTTGTCAAGAAATTTTATGCAAAATG 1614

QY  1340 TCATTGCAAGTATGAGAACATTTGACAGACTTCATGTTCTTGGACGGAAGCTATCTTC 1399
    || || || || || || || || || || || || || || || || || || || ||
DB  1615 TGAATGCTAGCATGCAACACTGTATGACTTTTCTTCTTGGAACAAAGCAAAATTTCTA 1674

QY  1400 TTGCTGCTGTGATGACTTTTGTGTCACGAGACCCCTCTGTGATGCCAAAGCTAGCT 1459
    || || || || || || || || || || || || || || || || || || || ||
DB  1675 TGGGAAGAGTTGGGAGTATTTTGTGTCACAGATCCAAATGATGATCAATGAGGAAATT 1734

QY  1460 TTTGGCTCCAAGAGTGCACATGTTGATGTCACAGACAGCTTGTGATGGGGAAGT 1519
    || || || || || || || || || || || || || || || || || || || ||
DB  1735 TTTGGTTCAGAGTGTGACATGATTCATTTGCTTACAAATAGTTTATGATGGGACAAA 1794

QY  1520 TCATCCACCTGCTGAGGATATGTTGCATCCACCCCTTGTGCGCTTCATGCTG 1579
    || || || || || || || || || || || || || || || || || || || ||
DB  1795 TGAATTCAGCTGATTTGGGATATGTTCCAAATCAGATCATGTTGTGCTAAAGTTGATGCTG 1854

QY  1580 CCTTCGAGCCATCTCTGCTGGCCCGCATCTATGTTAGTATCTGTGGAAAGCATAACT 1639
    || || || || || || || || || || || || || || || || || || || ||
DB  1855 GTTCAAGAGCTATTTGTGTTGGCCAAATCTATGTGATGATATGTTGCTCATGATTT 1914

QY  1640 TTGATCTTCTGAAGAAACTAGTGTCTCTGATGATGATCTCTTGAAGTGAAGTACTATG 1699

```

```

DB  1915 TTGATTTGATTAAGAGCTTTGATTCGCCGATGATGATACATACCAAAATGATATTTTTC 1974
    || || || || || || || || || || || || || || || || || || || ||
QY  1700 CACTCCCGCATCGCATTTGTTGTTGAAGACCCTTGGATATATGAGAAACTATAGCTTA 1759
    || || || || || || || || || || || || || || || || || || || ||
DB  1975 CATTTCACCAATCTGAGATGTCTTTTCAAAAACCTTTATGACATACACATGTCCTCA 2034
    || || || || || || || || || || || || || || || || || || || ||
QY  1760 AGATTGGATCTCAACAAAGTTCACGTGAGATGATGTCATTAACCTGCCAAGAGAG 1819
    || || || || || || || || || || || || || || || || || || || ||
DB  2035 AATTTTGGAACTTCAACCAAGTATGAGAGGCTGATTTGTCCTTCAACGTCTCAAGGGGCG 2094
    || || || || || || || || || || || || || || || || || || || ||
QY  1820 GATGTTGCTGAGACAGCGCCCAACCAATGCTTTTACATATCTCAAAAGCATGACAT 1879
    || || || || || || || || || || || || || || || || || || || ||
DB  2095 GGTGGATCTCAATTAATGCAAGAGTTTAGGGGCTTCCCTGAATGCTAACAAGCCAAATACCG 2154
    || || || || || || || || || || || || || || || || || || || ||
QY  1880 CCAAAACTAACCACAAAGCATAGAAATGCGACAGTGGAGAAAAACCTATCTATTTGAAG 1939
    || || || || || || || || || || || || || || || || || || || ||
DB  2155 GTACTGTTCTATGTAAACCGAAGTTGAATGGATGCAAAAGGAAAGAACATCTCATTTGGGTA 2214
    || || || || || || || || || || || || || || || || || || || ||
QY  1940 GCGTTAAACCTTTGCGCTTACCTCTATCAAGCCAAAAACT---TATCTCTCCAAAGC 1996
    || || || || || || || || || || || || || || || || || || || ||
DB  2215 AAGCAGAAACATATATGATGCTACTTGAATCAAGCTGAGAGCTTTCTTGTATGATCTTA 2274
    || || || || || || || || || || || || || || || || || || || ||
QY  1997 CCTTCACAGATCTTGACATAGCTCTTGACCCATTCGAATTCGAGCTCATCTGTTTAC 2056
    || || || || || || || || || || || || || || || || || || || ||
DB  2275 AATCCGAACCAATTCAGTTACTATTCACACATCTACATTTGAGCTATACAGTTTGTTC 2334
    || || || || || || || || || || || || || || || || || || || ||
QY  2057 CAGTACCAAACTCATCCAACTTCTTACACTTTGGCCCAATTTGGGCTGGTGAACATGC 2116
    || || || || || || || || || || || || || || || || || || || ||
DB  2335 CCGTTACAAA---GTTATGTGGCGGCATCAAAATTTGACCAAAATTTGGTGGACAAACATGT 2391
    || || || || || || || || || || || || || || || || || || || ||
QY  2117 TTAACACTAGTGAGACCATCATCTGTGACTATGACGATGACCTAAGCTAGTCTGAGAG 2176
    || || || || || || || || || || || || || || || || || || || ||
DB  2392 TCAACAGTGCTGAACAGATTATAGATTGGAAATGTTG-----GAATGCTGCTAAGA 2445
    || || || || || || || || || || || || || || || || || || || ||
QY  2177 TTGCTGTCAAGAGGTGTGTGATGCGAGTATTGTCATCGAAAAACCAAGGCTTGTG 2236
    || || || || || || || || || || || || || || || || || || || ||
DB  2446 TTAAGATTAAAGGTGTGTGAGATTCTTCTGCTTATTTAGTAAATCACCAAGAAAGTTTC 2505
    || || || || || || || || || || || || || || || || || || || ||
QY  2237 GTATTGATGGGAGAGATGTTGGTTCACAGTATGATCAGGACCAAAATGCTGTGCTCAG 2296
    || || || || || || || || || || || || || || || || || || || ||
DB  2506 AGTTGAATGTTGTGAGAGTGATTTTGAAGTGGCTAGTGATGGAATTTGTGTCAATG 2565
    || || || || || || || || || || || || || || || || || || || ||
QY  2297 TGCCATGCCCAATTTGATCTT 2317
    || || || || || || || || || || || || || || || || || || || ||
DB  2566 TTCTTGAATGAAGAGCCTT 2586
    || || || || || || || || || || || || || || || || || || || ||

```

Search completed: April 3, 2003, 18:19:27
 Job time : 7057 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

Run on: April 3, 2003, 12:29:40 ; Search time 592 Seconds
(without alignments)
8947.128 Million cell updates/sec

Title:	US-09-675-208-4_COPY_56_2407
Perfect score:	2352
Sequence:	1 ATGGCTCCAGTTTAAAAA.....CGTTATCGACTACTGTTT 2352

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

```
Searched:      2185239 seqs, 1125999159 residues
Total number of hits satisfying chosen parameters: 4370478
```

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing:	Minimum Match	0%
	Maximum Match	100%
	Listing first	45 summaries

Database : N_Geneseq_101002:1

1:	/SID52/gcgdata/genseq/genseqn-emb1/NA1960.DAT.*
2:	/SID52/gcgdata/genseq/genseqn-emb1/NA1961.DAT.*
3:	/SID52/gcgdata/genseq/genseqn-emb1/NA1982.DAT.*
4:	/SID52/gcgdata/genseq/genseqn-emb1/NA1993.DAT.*
5:	/SID52/gcgdata/genseq/genseqn-emb1/NA1984.DAT.*
6:	/SID52/gcgdata/genseq/genseqn-emb1/NA1965.DAT.*
7:	/SID52/gcgdata/genseq/genseqn-emb1/NA1966.DAT.*
8:	/SID52/gcgdata/genseq/genseqn-emb1/NA1967.DAT.*
9:	/SID52/gcgdata/genseq/genseqn-emb1/NA1968.DAT.*
10:	/SID52/gcgdata/genseq/genseqn-emb1/NA1989.DAT.*
11:	/SID52/gcgdata/genseq/genseqn-emb1/NA1990.DAT.*
12:	/SID52/gcgdata/genseq/genseqn-emb1/NA1991.DAT.*
13:	/SID52/gcgdata/genseq/genseqn-emb1/NA1992.DAT.*
14:	/SID52/gcgdata/genseq/genseqn-emb1/NA1993.DAT.*
15:	/SID52/gcgdata/genseq/genseqn-emb1/NA1994.DAT.*
16:	/SID52/gcgdata/genseq/genseqn-emb1/NA1995.DAT.*
17:	/SID52/gcgdata/genseq/genseqn-emb1/NA1996.DAT.*
18:	/SID52/gcgdata/genseq/genseqn-emb1/NA1997.DAT.*
19:	/SID52/gcgdata/genseq/genseqn-emb1/NA1998.DAT.*
20:	/SID52/gcgdata/genseq/genseqn-emb1/NA1999.DAT.*
21:	/SID52/gcgdata/genseq/genseqn-emb1/NA2000.DAT.*
22:	/SID52/gcgdata/genseq/genseqn-emb1/NA2001A.DAT.*
23:	/SID52/gcgdata/genseq/genseqn-emb1/NA2001B.DAT.*
24:	/SID52/gcgdata/genseq/genseqn-emb1/NA2002.DAT.*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
1	2352	100.0	2517	19	AAV22250		Cucumber raffinose
2	2352	100.0	2569	20	AAH61238		Cucumber raffinose
3	1041.4	44.3	2690	20	AAZ20209		Mustard raffinose
4	977	41.5	2498	19	AAV40801		Soybean raffinose
5	976.4	41.5	2498	22	AAH89523		Soybean raffinose
6	975.4	41.5	2690	20	AAZ20208		Sugarbeet raffinose
7	966.8	41.1	2343	22	AAH27438		Soybean polynucleo
8	966	41.1	2497	20	AAZ10002		Nucleic acid encoco
9	956.6	40.7	2746	19	AAV40800		Broad bean raffinoc

patience
Sgathelash

10	956.6	40.7	2746	20	AAZ100031	Nucleic acid encod
11	948.8	40.3	2524	21	AAAD00335	Soybean raffinose
12	886.8	37.7	1762	19	AAV40802	Japanese arlichoke
13	886.8	37.7	1762	20	AAZ100030	Nucleic acid encod
14	866.4	36.8	1762	20	AAZ20210	Rapeseed raffinose
15	303.8	12.9	928	20	AAZ20207	Soybean raffinose
16	275.8	11.7	2780	20	AAK61239	Soybean raffinose
17	243.2	10.3	2868	21	AAAD00332	Corn raffinose syn
18	240.8	10.2	2842	21	AAAD00334	Corn raffinose syn
19	220.8	9.4	2853	21	AAAD00338	Wheat raffinose syn
20	191.8	8.2	2668	21	AAAD00337	Wheat raffinose syn
21	184	7.8	3060	21	AAAD00333	Rice raffinose syn
22	174.4	7.4	1816	21	AAAD00330	Corn raffinose syn
23	169.6	7.2	993	19	AAV40803	Corn raffinose syn
24	169.6	7.2	993	20	AAZ10004	Nucleic acid encod
25	114	4.8	822	21	AAAC42001	Arabidopsis thalia
26	113.2	4.8	1036	24	AAAN96412	Arabidopsis thalia
27	90.8	3.9	293	20	AAK61225	Raffinose synthase
28	75.4	3.2	296	24	ABL73625	Corn tassal-derivate
29	50.4	2.1	636	22	AAH87794	Peppermint plant c
30	48.2	2.0	540	22	AAH87795	Peppermint plant c
31	41.6	1.8	29255	23	AAAS58516	Protonibacterium
32	39.4	1.7	149671	24	ABK84797	Human cDNA differe
33	39.2	1.7	1417	24	ABK11710	DNA encoding novel
34	39.2	1.7	3408	22	AAI93894	Human stomach cance
35	39.2	1.7	3408	22	AAH18536	Human cDNA sequenc
36	39.2	1.7	66566	21	AAVA53450	Human chlorodoxin
37	38.8	1.6	189	24	AAEN70953	Human ORF4040 CDNA
38	38.8	1.6	4590	22	AAAH24065	Yeast ADD9604-8sscc
39	38.2	1.6	28046	24	ABK65837	Human secreted prote
40	38	1.6	911	22	AAAD03342	Probe 11 for screed
41	38	1.6	911	22	AAAD03343	Probe 11 for screed
42	37.8	1.6	31422	21	AAA923302	S. avernifilis avern
43	37.8	1.6	31422	22	AAH79278	Streptomyces avern
44	37.4	1.6	1664	22	AAAS35771	Human cardiovascu
45	37.4	1.6	1664	22	AAAS35772	Human cardiovascu

ALIGNMENTS

RESULT 1
AAV22250
ID AAV22250 standard; cDNA to mRNA; 2517 BP

AC AAV22250;

DT 06-JUL-1998 (first entry)

Cucumber raffinose synthase cDNA.

KW Cucumber; raffinose synthase; sucrose; galactinol; ds

OS Cucumis sativus

Key	Location/Qualifiers
FH	
EC	

```

/*tag= a
/next= a

```

XX TP10094073-4

XX 07-ANB-1009
DD

XX 07TD-0111134
DE 38-ADP-1007

XX
DB 36-TIT-1006. 06TB-0108070

PR 26-APR-1996; 96JP-010/682;
XX

PA (AJIN) AJINOMOTO KK.

DR WPI; 1998-264858/24.
D-DEDP; AWE3E70

XX Raffinose synthase gene - useful for preparation of raffinose in
PT transformed plant
XX
XX
PS Claim 7, Pages 17-20; 26pp; Japanese.
XX
CC The present sequence encodes cucumber raffinose synthase, which
CC forms raffinose from sucrose and galactinol, has an optimum pH of
CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
CC iodoacetamide, N-ethylmaleimide and myo-inositol.
XX
SQ Sequence 2517 BP; 613 A; 524 C; 642 G; 738 T; 0 other;
Query Match 100.0%; Score 2352; DB 19; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGGCTCTAGTTTAAATAAGTGCTGCCAACGATGTTTCATTTGATGGCTTAAATGAC 60
DB ATGGCTCTAGTTTAAATAAGTGCTGCCAACGATGTTTCATTTGATGGCTTAAATGAC 115
QY 61 ATGTGCTACCGTTGTCANCGAGATGCGATTCACGTGAGAGGTCATTCGTTTCG 120
DB ATGTGCTACCGTTGTCANCGAGATGCGATTCACGTGAGAGGTCATTCGTTTCG 175
QY 121 TCCGATGTTCTGAGAACATTTGCTTCCTTCCTACACTTCGATAGCAAGTCC 180
DB TCCGATGTTCTGAGAACATTTGCTTCCTTCCTACACTTCGATAGCAAGTCC 235
QY 176 TCGAGATTCCTGAGAACATTTGCTTCCTTCCTACACTTCGATAGCAAGTCC 235
QY 181 CCGGTTTCGGTTGGTTGCTTTGTTGGATTCGACGCGTGGAACTGATAGCCGACATGTT 240
DB CCGGTTTCGGTTGGTTGCTTTGTTGGATTCGACGCGTGGAACTGATAGCCGACATGTT 295
QY 241 GTTTCGATGGGAAGCTGAGAGATATTCGTTTATGAGATTTTTCAGGTTTAAAGTTTG 300
DB GTTTCGATGGGAAGCTGAGAGATATTCGTTTATGAGATTTTTCAGGTTTAAAGTTTG 355
QY 301 TGGACTACACACTGGGTTGGTGGAAATGGTGGGATCTTGAATCGAGAGCTCAGATTGTG 360
DB TGGACTACACACTGGGTTGGTGGAAATGGTGGGATCTTGAATCGAGAGCTCAGATTGTG 415
QY 356 TGGACTACACACTGGGTTGGTGGAAATGGTGGGATCTTGAATCGAGAGCTCAGATTGTG 415
QY 361 ATCCTTGAGAAAGTCAGATTCTGGTGCAGCGTATGTTTCTCTTCGATCGTTGAGGGA 420
DB ATCCTTGAGAAAGTCAGATTCTGGTGCAGCGTATGTTTCTCTTCGATCGTTGAGGGA 475
QY 416 ATCCTTGAGAAAGTCAGATTCTGGTGCAGCGTATGTTTCTCTTCGATCGTTGAGGGA 475
DB ATCCTTGAGAAAGTCAGATTCTGGTGCAGCGTATGTTTCTCTTCGATCGTTGAGGGA 480
QY 421 CCGTTCCGAACCTCGATTCAGCTGGGATGATGACTTTGTGATGTTTGTGCGAGAT 480
DB CCGTTCCGAACCTCGATTCAGCTGGGATGATGACTTTGTGATGTTTGTGCGAGAT 535
QY 476 CCGTTCCGAACCTCGATTCAGCTGGGATGATGACTTTGTGATGTTTGTGCGAGAT 535
DB CCGTTCCGAACCTCGATTCAGCTGGGATGATGACTTTGTGATGTTTGTGCGAGAT 540
QY 481 GGTTCGTCGAAGATGTTGATGATCGATTCGGAAGTATGTTTATCTTCATGCTGGTGTAT 540
DB GGTTCGTCGAAGATGTTGATGATCGATTCGGAAGTATGTTTATCTTCATGCTGGTGTAT 595
QY 536 GGTTCGTCGAAGATGTTGATGATCGATTCGGAAGTATGTTTATCTTCATGCTGGTGTAT 595
DB GGTTCGTCGAAGATGTTGATGATCGATTCGGAAGTATGTTTATCTTCATGCTGGTGTAT 600
QY 541 GATTCGTTTGCACTTGTGTAAGAGGCGATGAAGATCGTAGAGACCATCTTGGAACTTTT 600
DB GATTCGTTTGCACTTGTGTAAGAGGCGATGAAGATCGTAGAGACCATCTTGGAACTTTT 655
QY 596 GATTCGTTTGCACTTGTGTAAGAGGCGATGAAGATCGTAGAGACCATCTTGGAACTTTT 655
DB GATTCGTTTGCACTTGTGTAAGAGGCGATGAAGATCGTAGAGACCATCTTGGAACTTTT 660
QY 601 GCCTTGTGAGAGAGAGCTCCACAGGTATCGTGGACAATTTGGTGGTGGACAGTGG 660
DB GCCTTGTGAGAGAGAGCTCCACAGGTATCGTGGACAATTTGGTGGTGGACAGTGG 715
QY 656 GCCTTGTGAGAGAGAGCTCCACAGGTATCGTGGACAATTTGGTGGTGGACAGTGG 715
DB GCCTTGTGAGAGAGAGCTCCACAGGTATCGTGGACAATTTGGTGGTGGACAGTGG 720
QY 661 GACGGCTTTTACCTAACGCTTCATCCACAGGCGTAAATGAGAGCGTGGAGATCTCGTC 720
DB GACGGCTTTTACCTAACGCTTCATCCACAGGCGTAAATGAGAGCGTGGAGATCTCGTC 775
QY 716 GACGGCTTTTACCTAACGCTTCATCCACAGGCGTAAATGAGAGCGTGGAGATCTCGTC 775
DB GACGGCTTTTACCTAACGCTTCATCCACAGGCGTAAATGAGAGCGTGGAGATCTCGTC 780
QY 721 GACGGCGGTTTGTCTCCCGGTTTAACTCTTAATCGACGATGTTTGGCAATCCATCGACAC 780
DB GACGGCGGTTTGTCTCCCGGTTTAACTCTTAATCGACGATGTTTGGCAATCCATCGACAC 835
QY 776 GACGGCGGTTTGTCTCCCGGTTTAACTCTTAATCGACGATGTTTGGCAATCCATCGACAC 835
DB GACGGCGGTTTGTCTCCCGGTTTAACTCTTAATCGACGATGTTTGGCAATCCATCGACAC 840
QY 781 GATTGGGATCCCATCAACCAAGAGATGAACCAACCGTGGCGGACGACAATGCCC 840
DB GATTGGGATCCCATCAACCAAGAGATGAACCAACCGTGGCGGACGACAATGCCC 840

DB 836 GATTGGGATCCCATCAACCAAGAGATGAACCAACCGTGGCGGACGACAATGCCC 895
QY 841 TGGCGTCTTTTGAATTTCCAGAGAAATTCGAATTCCTGGATAGCAATCCCAAGGCC 900
DB TGGCGTCTTTTGAATTTCCAGAGAAATTCGAATTCCTGGATAGCAATCCCAAGGCC 955
QY 901 ACCGGCCCCGAGCCGCGCAGAGAAAGGAGATGAAAGCGTTTATAGATGACTCAAGAGAG 960
DB ACCGGCCCCGAGCCGCGCAGAGAAAGGAGATGAAAGCGTTTATAGATGACTCAAGAGAG 1015
QY 956 ACCGGCCCCGAGCCGCGCAGAGAAAGGAGATGAAAGCGTTTATAGATGACTCAAGAGAG 1015
QY 961 TTTTAACAGCTGAGACATGTTTATGTTGGCATGCTTTGTGTGATATTTGGGTGGCTTT 1020
DB TTTTAACAGCTGAGACATGTTTATGTTGGCATGCTTTGTGTGATATTTGGGTGGCTTT 1075
QY 1016 TTTTAACAGCTGAGACATGTTTATGTTGGCATGCTTTGTGTGATATTTGGGTGGCTTT 1075
DB TTTTAACAGCTGAGACATGTTTATGTTGGCATGCTTTGTGTGATATTTGGGTGGCTTT 1105
QY 1021 CCGCCCGCAGGTGCTGCTGCTGCTGAGGACGCTGATTCAGCAGCAGTTCCTCACCAGG 1080
DB CCGCCCGCAGGTGCTGCTGCTGCTGAGGACGCTGATTCAGCAGCAGTTCCTCACCAGG 1135
QY 1076 CCGCCCGCAGGTGCTGCTGCTGCTGAGGACGCTGATTCAGCAGCAGTTCCTCACCAGG 1135
DB CCGCCCGCAGGTGCTGCTGCTGCTGAGGACGCTGATTCAGCAGCAGTTCCTCACCAGG 1140
QY 1081 CTGACAGATGACGATGAGAGATTTTGGCGGTGATTAAGATTGTTCTTCATTAAGTGGGCTG 1140
DB CTGACAGATGACGATGAGAGATTTTGGCGGTGATTAAGATTGTTCTTCATTAAGTGGGCTG 1195
QY 1136 CTGACAGATGACGATGAGAGATTTTGGCGGTGATTAAGATTGTTCTTCATTAAGTGGGCTG 1195
DB CTGACAGATGACGATGAGAGATTTTGGCGGTGATTAAGATTGTTCTTCATTAAGTGGGCTG 1200
QY 1141 CTGCGCGCGGAGAGAGGCTGAGGAGATGTACGAAGCACTTCATGCTCATTTGGAAAAAGTT 1200
DB CTGCGCGCGGAGAGAGGCTGAGGAGATGTACGAAGCACTTCATGCTCATTTGGAAAAAGTT 1255
QY 1196 GTCCCGCCGAGAGAGGCTGAGGAGATGTACGAAGCACTTCATGCTCATTTGGAAAAAGTT 1255
DB GTCCCGCCGAGAGAGGCTGAGGAGATGTACGAAGCACTTCATGCTCATTTGGAAAAAGTT 1260
QY 1201 GGGATCGACGAGTGTAAAGATTGACCTTATCCACTTATGGAGATGTTGTGTAAGACTAT 1260
DB GGGATCGACGAGTGTAAAGATTGACCTTATCCACTTATGGAGATGTTGTGTAAGACTAT 1315
QY 1256 GGGATCGACGAGTGTAAAGATTGACCTTATCCACTTATGGAGATGTTGTGTAAGACTAT 1315
DB GGGATCGACGAGTGTAAAGATTGACCTTATCCACTTATGGAGATGTTGTGTAAGACTAT 1320
QY 1261 GGAAGGAGAGTGTGATTTGGCAAAAGCATATTACAAAGCAATGACCAATCAATTAATAA 1320
DB GGAAGGAGAGTGTGATTTGGCAAAAGCATATTACAAAGCAATGACCAATCAATTAATAA 1375
QY 1316 GGAAGGAGAGTGTGATTTGGCAAAAGCATATTACAAAGCAATGACCAATCAATTAATAA 1375
DB GGAAGGAGAGTGTGATTTGGCAAAAGCATATTACAAAGCAATGACCAATCAATTAATAA 1380
QY 1321 CATTTTAAAGAAATGAGATGATTCGAAATGGAATTTGAAAGCACTTCATGTTCCCTT 1380
DB CATTTTAAAGAAATGAGATGATTCGAAATGGAATTTGAAAGCACTTCATGTTCCCTT 1435
QY 1376 CATTTTAAAGAAATGAGATGATTCGAAATGGAATTTGAAAGCACTTCATGTTCCCTT 1435
DB CATTTTAAAGAAATGAGATGATTCGAAATGGAATTTGAAAGCACTTCATGTTCCCTT 1440
QY 1381 GGCACGAGAGCTATCTCTCTGCTGCTGTTGGTGAATGATGATTTGGTGCAGAGACCCCTT 1440
DB GGCACGAGAGCTATCTCTCTGCTGCTGTTGGTGAATGATGATTTGGTGCAGAGACCCCTT 1495
QY 1436 GGCACGAGAGCTATCTCTCTGCTGCTGTTGGTGAATGATGATTTGGTGCAGAGACCCCTT 1495
DB GGCACGAGAGCTATCTCTCTGCTGCTGTTGGTGAATGATGATTTGGTGCAGAGACCCCTT 1500
QY 1441 GGTGATCCAAAGGATACGTTTGGCTCCAAAGATGTCACATGTTTATTTGTCACAGAC 1500
DB GGTGATCCAAAGGATACGTTTGGCTCCAAAGATGTCACATGTTTATTTGTCACAGAC 1555
QY 1496 GGTGATCCAAAGGATACGTTTGGCTCCAAAGATGTCACATGTTTATTTGTCACAGAC 1555
DB GGTGATCCAAAGGATACGTTTGGCTCCAAAGATGTCACATGTTTATTTGTCACAGAC 1560
QY 1501 AGCTTGTGATGGGAGACTTCATCCACCCTGAGTGGAGATGTTTCCAAATCCACCACCTT 1560
DB AGCTTGTGATGGGAGACTTCATCCACCCTGAGTGGAGATGTTTCCAAATCCACCACCTT 1615
QY 1556 AGCTTGTGATGGGAGACTTCATCCACCCTGAGTGGAGATGTTTCCAAATCCACCACCTT 1615
DB AGCTTGTGATGGGAGACTTCATCCACCCTGAGTGGAGATGTTTCCAAATCCACCACCTT 1620
QY 1561 TGTGCCGCTTCATGCTGCTCTCGAGGCAATCTGTGTGGCCCGATCTATGTTAGTAT 1620
DB TGTGCCGCTTCATGCTGCTCTCGAGGCAATCTGTGTGGCCCGATCTATGTTAGTAT 1675
QY 1616 TGTGCCGCTTCATGCTGCTCTCGAGGCAATCTGTGTGGCCCGATCTATGTTAGTAT 1675
DB TGTGCCGCTTCATGCTGCTCTCGAGGCAATCTGTGTGGCCCGATCTATGTTAGTAT 1680
QY 1621 TGTGTGGGAAAGCATTAATTTGATCTTCTGAAAAAACTAGTGTCTCTATGATGATG 1680
DB TGTGTGGGAAAGCATTAATTTGATCTTCTGAAAAAACTAGTGTCTCTCTATGATGATG 1735
QY 1676 TGTGTGGGAAAGCATTAATTTGATCTTCTGAAAAAACTAGTGTCTCTCTATGATGATG 1735
DB TGTGTGGGAAAGCATTAATTTGATCTTCTGAAAAAACTAGTGTCTCTCTCTATGATGATG 1740
QY 1681 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
DB CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795
QY 1736 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795
DB CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
QY 1741 AATGGAAGAACTATGCTTAAGATTTGGAATTCACCAAGTTCATGAGATGATGATGATGATGAT 1800
DB AATGGAAGAACTATGCTTAAGATTTGGAATTCACCAAGTTCATGAGATGATGATGATGATGAT 1855
QY 1796 AATGGAAGAACTATGCTTAAGATTTGGAATTCACCAAGTTCATGAGATGATGATGATGAT 1855
DB AATGGAAGAACTATGCTTAAGATTTGGAATTCACCAAGTTCATGAGATGATGATGATGAT 1860
QY 1801 TTCAACTGCAAGAGAGAGATGTTGCTGAGACAGCCGCAACCAATGCTTTTCACAA 1860
DB TTCAACTGCAAGAGAGAGATGTTGCTGAGACAGCCGCAACCAATGCTTTTCACAA 1915
QY 1856 TTCAACTGCAAGAGAGAGATGTTGCTGAGACAGCCGCAACCAATGCTTTTCACAA 1915
DB TTCAACTGCAAGAGAGAGATGTTGCTGAGACAGCCGCAACCAATGCTTTTCACAA 1920
QY 1861 TTACTCAAAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
DB TTACTCAAAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1975
QY 1916 TTACTCAAAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1975
DB TTACTCAAAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1975

Oy 1921 AACCTATCTCTATTGAAAGGCGTTAAACCTTTGGCTTACCTTATCAAGCCAAAAA 1980
|||||
Db 1976 AACCTATCTCTATTGAAAGGCGTTAAACCTTTGGCTTACCTTATCAAGCCAAAAA 2035
Oy 1981 CTTATCTCTCCAAAGCCCTCTCAGATCTTGACATAGCTCTTGACCCATTCGATTCGAG 2040
|||||
Db 2036 CTTATCTCTCCAAAGCCCTCTCAGATCTTGACATAGCTCTTGACCCATTCGATTCGAG 2095
Oy 2041 CTCATCAGCTGTTCCAGTCAGTCAGCAACCTCTCTACCTTCCCTCCCAATT 2100
|||||
Db 2096 CTCATCAGCTGTTCCAGTCAGTCAGCAACCTCTCTACCTTCCCTCCCAATT 2155
Oy 2101 GGGCTGTGCAACATGCTTAACTAGTGGAGCCATCAATCTGTGAGCATGAGCATGAC 2160
2156 GGGCTGTGCAACATGCTTAACTAGTGGAGCCATCAATCTGTGAGCATGAGCATGAC 2215
Oy 2161 CTAAGCTCAGTCGAGATTTGGTGTCAAAAGGGTGGTGAAGATCGAGTATTTGCATCGAAA 2220
2216 CTAAGCTCAGTCGAGATTTGGTGTCAAAAGGGTGGTGAAGATCGAGTATTTGCATCGAAA 2275
Oy 2221 AAGCAAGGCGCTTGTCTGATTCATGAGGAGATGTTGGTTCAATGATGATGAGCAACAA 2280
2276 AAGCAAGGCGCTTGTCTGATTCATGAGGAGATGTTGGTTCAATGATGATGAGCAACAA 2335
Oy 2281 ATGCTGTGTGTTCAAGTCGATGCGCAATTGATCTTCTCGGTGCGATTTCCGTTATC 2340
2336 ATGCTGTGTGTTCAAGTCGATGCGCAATTGATCTTCTCGGTGCGATTTCCGTTATC 2395
Oy 2341 GAGTACTTGTGTT 2352
2396 GAGTACTTGTGTT 2407
Db

RESULT 2
AAK61238
ID AAK61238 standard: cDNA to mRNA: 2569 BP.

XX AAK61238:
XX
XX 29-JUL-1999 (first entry)
XX
XX Cucurbit raffinose synthase encoding cDNA.
XX DE
XX Raffinose synthase; sucrose; galactinol; ds.
XX KM
XX Cucumis sativus.
XX OS

XX
XX Key Location/Qualifiers
XX CDS 56..2410
XX /*tag= a

XX JPI1123080-A.

XX 11-MAY-1999.

XX 24-OCT-1997; 97JP-0292969.

XX 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

XX MPI: 1999-340516/29.

XX P-PSDB: AAV17417.

XX New raffinose synthase gene - for production of raffinose from
XX sucrose and galactinol

XX Claim 10; Page 22-25; 37pp; Japanese.

XX The present invention describes a raffinose synthase, having an activity
XX of forming raffinose from sucrose and galactinol. The raffinose synthase
XX gene can be used for expression in a plant for the production of

CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence encodes raffinose
XX synthase from cucumber.

XX Sequence 2569 BP: 641 A; 527 C; 643 G; 755 T; 3 other:

Query Match 100.0%; Score 2352; DB 20; Length 2569;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 ATGGCTCTCAGTATTAAAAATGGTGGCTTCACAGTATGTTTCATTTGATGCGTTAAATGAC 60
Db 56 ATGGCTCTCAGTATTAAAAATGGTGGCTTCACAGTATGTTTCATTTGATGCGTTAAATGAC 115
Oy 61 ATGTCGTACCGTTTGGCATGAGGAGATCGGATTCGATTCAGTGAAGGATTCGTTGCG 120
Db 116 ATGTCGTACCGTTTGGCATGAGGAGATTCGATTCAGTGAAGGATTCGTTGCG 175
Oy 121 TCCGATGTTCCAGAGACATTTGCTTCTCTCTCCGTACCTTCGATACATGACAAAGTCC 180
Db 176 TCCGATGTTCCAGAGACATTTGCTTCTCTCTCCGTACCTTCGATACATGACAAAGTCC 235
Oy 181 CCGGTTTCCGTTGGTTCCTTTGTTGTTGATTCGACGCGTGGAACTGATAGCCGACATGTT 240
Db 236 CCGGTTTCCGTTGGTTCCTTTGTTGTTGATTCGACGCGTGGAACTGATAGCCGACATGTT 295
Oy 241 GTTTCGATTTGGGAGACCTGAGATATTCGGTTATGAGATTTTACGATTTAAAGGTTTGG 300
Db 296 GTTTCGATTTGGGAGACCTGAGATATTCGGTTATGAGATTTTACGATTTAAAGGTTTGG 355
Oy 301 TGGACTACACACTGGGTTTGGTTCGAATGGTGGGATCTTGAATCGGAGACTCAGATTGTG 360
Db 356 TGGACTACACACTGGGTTTGGTTCGAATGGTGGGATCTTGAATCGGAGACTCAGATTGTG 415
Oy 361 ATCTTTGAGAACTCAGATTTCTGCTGACCGCTATGTTTTCCTTCCGATCGTTGAGGA 420
Db 416 ATCTTTGAGAACTCAGATTTCTGCTGACCGCTATGTTTTCCTTCCGATCGTTGAGGA 475
Oy 421 CCGTTCCGAACTGATTCAGCTGGGATGATGATGATTTGTCGATGTTTGGTTCGAGAGT 480
Db 476 CCGTTCCGAACTGATTCAGCTGGGATGATGATGATTTGTCGATGTTTGGTTCGAGAGT 535
Oy 481 GGTTCGTGAAAAGTTGTTGATGATCATGTTCCGAAATGTTGATCTTCATCTGCTGTAT 540
Db 536 GGTTCGTGAAAAGTTGTTGATGATCATGTTCCGAAATGTTGATCTTCATCTGCTGTAT 595
Oy 541 GATCCGTTTGCACCTTTGTTAAAGAGCGATGAAGATCTGTGAGAGCCCATCTTGAACCTTTT 600
Db 596 GATCCGTTTGCACCTTTGTTAAAGAGCGATGAAGATCTGTGAGAGCCCATCTTGAACCTTTT 655
Oy 601 CGCTTGTGGAGAGAAAGCTCCACAGGTATGCTGAGCAAAATTTGGTTCGTCACAGTCG 660
Db 656 CGCTTGTGGAGAGAAAGCTCCACAGGTATGCTGAGCAAAATTTGGTTCGTCACAGTCG 715
Oy 661 GACGCGTTTACCTTAACGCTTCATCCACAGGCGTAAATGAAGGCGTGAAGCATCTCGTC 720
Db 716 GACGCGTTTACCTTAACGCTTCATCCACAGGCGTAAATGAAGGCGTGAAGCATCTCGTC 775
Oy 721 GACGCGGTTTGTCTCCGCGTTTAACTCTAATTCAGAGATGTTGGCAATCCATCCGACAC 780
Db 776 GACGCGGTTTGTCTCCGCGTTTAACTCTAATTCAGAGATGTTGGCAATCCATCCGACAC 835
Oy 781 GATTCGGAGTCCCATCAGCAAGAAAGAAATGAGCAACCGTCCGCGGAGCAAAATGCC 840
Db 836 GATTCGGAGTCCCATCAGCAAGAAAGAAATGAGCAACCGTCCGCGGAGCAAAATGCC 895
Oy 841 TCCGCTCTTTTGAATTCCAAGAGATTTACAATTCCTGATACGTAAATCCCAAGGCC 900
Db 896 TCCGCTCTTTTGAATTCCAAGAGATTTACAATTCCTGATACGTAAATCCCAAGGCC 955
Oy 901 ACCGCGCCCGGAGCGCCAGAGGAGATGAAGGCGTTTATGATGAACCTCAAGAGAG 960
Db 956 ACCGCGCCCGGAGCGCCAGAGGAGATGAAGGCGTTTATGATGAACCTCAAGAGAG 1015

QY 961 TTTAAGACTGTGGAGCATGTTATGTTGGCATGCTTTGTGGATATTGGGGTGGCCTT 1020
 |||||||
 Db 1016 TTTAAGACTGTGGAGCATGTTATGTTGGCATGCTTTGTGGATATTGGGGTGGCCTT 1075
 QY 1021 GGGCCCGAGTGGCTGGCTTGGCTGAGCAGTGTGATTCAGCCAGTGGCTTTCACAGGG 1080
 |||||||
 Db 1076 GGGCCCGAGTGGCTGGCTTGGCTGAGCAGTGTGATTCAGCCAGTGGCTTTCACAGGG 1135
 QY 1081 CTGCAGATGACGATGAGGATTTGGCGGTGATTAAGATTCTTCTATAGGTCGGCTG 1140
 |||||||
 Db 1136 CTGCAGATGACGATGAGGATTTGGCGGTGATTAAGATTCTTCTATAGGTCGGCTG 1195
 QY 1141 GTCCCGCCGAGAGGCTGAGAGATGTACGAAGACTTCATGCTCATTTGGAAAAAGTT 1200
 |||||||
 Db 1196 GTCCCGCCGAGAGGCTGAGGATGTACGAAGACTTCATGCTCATTTGGAAAAAGTT 1255
 QY 1201 GGGATCGAGGGTAAAGTTGACGTTATCCACTTTGGAGATGTTGTGTAAGCTAT 1260
 |||||||
 Db 1256 GGGATCGAGGGTAAAGTTGACGTTATCCACTTTGGAGATGTTGTGTAAGCTAT 1315
 QY 1261 GGAGGAGAGTGGATTTGGCAAGGCAATTTACAAAGCAATGACCAATCAATAATAA 1320
 |||||||
 Db 1316 GGAGGAGAGTGGATTTGGCAAGGCAATTTACAAAGCAATGACCAATCAATAATAA 1375
 QY 1321 CATTTTAAAGAAATGAGATCATTTGCAATGTGAACATTTGACAGCTTCATGTTCTT 1380
 |||||||
 Db 1376 CATTTTAAAGAAATGAGATCATTTGCAATGTGAACATTTGACAGCTTCATGTTCTT 1435
 QY 1381 GGCAGGAGAGTATCTCTCTGGTGGTGGTGGATGATCTTTGGTGCAGGAGCCCTCT 1440
 |||||||
 Db 1436 GGCAGGAGAGTATCTCTCTGGTGGTGGTGGATGATCTTTGGTGCAGGAGCCCTCT 1495
 QY 1441 GGTGATCCAAAGGATAGTTTTGGCTCAAGAGATGCACATGGTTCATTTGGCCAAGAC 1500
 |||||||
 Db 1496 GGTGATCCAAAGGATAGTTTTGGCTCAAGAGATGCACATGGTTCATTTGGCCAAGAC 1555
 QY 1501 AGCTGTGATGGGGAGACTTCATCCACCCTGACTGGGATATGTTCCAAATCCACCCT 1560
 |||||||
 Db 1556 AGCTGTGATGGGGAGACTTCATCCACCCTGACTGGGATATGTTCCAAATCCACCCT 1615
 QY 1561 TGTGCGGCTTCAGTCTCTCTGAGGCAATCTGTGGCCGCAATATGTTAGTATGAT 1620
 |||||||
 Db 1616 TGTGCGGCTTCAGTCTCTCTGAGGCAATCTGTGGCCGCAATATGTTAGTATGAT 1675
 QY 1621 TCTGTGGGAAGCATAACTTGTATCTTGAAAAAACTAGTGCCTCTGATGATGATC 1680
 |||||||
 Db 1676 TCTGTGGGAAGCATAACTTGTATCTTGAAAAAACTAGTGCCTCTGATGATGATC 1735
 QY 1681 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1740
 |||||||
 Db 1736 CTTCGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1795
 QY 1741 AATGGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1800
 |||||||
 Db 1796 AATGGAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1855
 QY 1801 TTCAACTGCGAAGGAGGAGATGTTGCTGAGACACGCCGCAACCAATGCTTTTCACAA 1860
 |||||||
 Db 1856 TTCAACTGCGAAGGAGGAGATGTTGCTGAGACACGCCGCAACCAATGCTTTTCACAA 1915
 QY 1861 TACTCAAAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1920
 |||||||
 Db 1916 TACTCAAAAAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1975
 QY 1921 AACCTATCTATTTGAAGGCGTTAAACCTTTGGCTTACTATCAAGCAACCAAAAA 1980
 |||||||
 Db 1976 AACCTATCTATTTGAAGGCGTTAAACCTTTGGCTTACTATCAAGCAACCAAAAA 2035
 QY 1981 CTTATCTCTCCAGCCCTCTCAAGATCTTGACATGCTTTGACCCATTCGAATTCGAG 2040
 |||||||
 Db 2036 CTTATCTCTCCAGCCCTCTCAAGATCTTGACATGCTTTGACCCATTCGAATTCGAG 2095

QY 2041 CTCATCAGTGTTCACCACTGACCAACATCATCCAAACTTCTCTACACTTTGCCCAATT 2100
 |||||||
 Db 2096 CTCATCAGTGTTCACCACTGACCAACATCATCCAAACTTCTCTACACTTTGCCCAATT 2155
 QY 2101 GGGCTGTGAACATGCTTAACTAGTGGAGCATCCATCTGTGGCATGACATGAC 2160
 |||||||
 Db 2156 GGGCTGTGAACATGCTTAACTAGTGGAGCATCCATCTGTGGCATGACATGAC 2215
 QY 2161 CTAACTCAGTGAATGTTGTGTCAAAGGGTGTGAGATGAGATTTTCATCGAATA 2220
 |||||||
 Db 2216 CTAACTCAGTGAATGTTGTGTCAAAGGGTGTGAGATGAGATTTTCATCGAATA 2275
 QY 2221 AAACCAAGGCTTGTGATGATGAGGAGATGTTGGSTTCAAGTATATGACAGCA 2280
 |||||||
 Db 2276 AAACCAAGGCTTGTGATGATGAGGAGATGTTGGSTTCAAGTATATGACAGCA 2335
 QY 2281 ATGGTGTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 2340
 |||||||
 Db 2336 ATGGTGTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCAT 2395
 QY 2341 GAGTACTGTTT 2352
 |||||||
 Db 2396 GAGTACTGTTT 2407

RESULT 3
 AA220209
 ID AA220209 standard; cDNA; 2690 BP.
 XX
 AC AA220209;
 DT 17-JAN-2000 (first entry)
 DE Mustard raffinose synthase cDNA.
 DE
 DE Mustard raffinose synthase cDNA.
 DE
 DE Mustard raffinose synthase; mustard; transgenic plant; ss.
 KW
 OS Brassica juncea.
 OS
 FT Key Location/Qualifiers
 FT CDS 134..2467
 FT /tag= a
 FT /note= "this region is specifically claimed in
 FT Claim 7"
 XX
 PN EP953643-A2.
 PD 03-NOV-1999.
 XX
 PR 27-APR-1999; 99EP-0107430.
 XX
 PR 30-APR-1998; 98JP-0120550.
 PR 30-APR-1998; 98JP-0120551.
 PR 04-DEC-1998; 98JP-0345590.
 PR 10-DEC-1998; 98JP-0351246.
 XX
 PA (SOMO) SUMITOMO CHEM CO LTD.
 XX
 PI Watanabe E, Oeda K;
 XX
 DR WPI: 1999-593144/51.
 DR P-PSDB: AAY32074.
 PT
 PT New sense and antisense genes, useful for altering the level of
 PT raffinose in food plants -
 PS
 PS Claim 7; Page 32-36; 55pp; English.
 CC This is the nucleotide sequence of a mustard cDNA clone coding
 CC for raffinose synthase (see AAY32074), a protein which can bind a
 CC D-galactosyl group through an alpha(1-6) bond to the hydroxy group
 CC attached to the carbon atom at the 6-position of the D-glucose
 CC residue in a sucrose molecule to form raffinose. The cDNA was

CC Isolated from mustard (Brassica juncea) leaf cDNA by PCR. Probes
CC or primers generated from plant raffinose synthase genes (see
CC AA20207-10) may be used to obtain other raffinose synthase genes
CC by labeled detection or amplification (claimed). These genes may
CC be used to control the levels of raffinose produced in plants.
CC Antisense genes can be used to knock out existing gene activity.
CC and sense genes to increase the level of gene activity. The
CC resulting transgenic plants may be used as a food source to alter
CC the growing conditions for gut enterobacteria, providing general
CC health advantages. The sense and antisense genes may also be used
CC in gene and phenotypic analysis of plants and for the selection of
CC plant clones with the desired characteristics with respect to
CC raffinose content.

SO Sequence 2690 BP: 648 A; 608 C; 668 G; 765 T; 1 other:

Query Match 44.3%; Score 1041.4; DB 20; Length 2690;
Best Local Similarity 68.3%; Pred. No. 0;
Matches 1530; Conservative 0; Mismatches 676; Indels 33; Gaps 5;

QY 80 TCAGAGGATCGGATTTCACTGTGAACGGTCATTGCTTCTGTCGGATGCTCGGAACA 139
DB 210 TAGAGGATTCGATCTCCATGCAATGGTACAGTTGTTAAACCGATGTGACGGTTAAACG 269
QY 140 TTGTTGCTTCTCTCTCCGTAC-----ACTTCGATAGACAAGTCCCGGTTTCG 189
DB 270 TCAGCTGTCACTGCTTCACTTACCTAGTACGACAAAGACGAGACCGGTTGAACGGCTCCG 329
QY 190 GTTGGTGTGTTGTGGATTCACACGCTGC--GAACCTATAGCCGACATGTTGTTTCA 247
DB 330 CTGGTTCATTTCATCGGGTTTAATCTCGACGGTGAGCCAGCAAGACCGCCGCTGCTCA 389
QY 248 TTGGGAAGCTGAAGATTTACGTTTATGATATTTTACGTTTAAAGTTTGATGACTA 307
DB 390 TCGGTAAGTCAAGGATATTCATTCATGACATATTCGTTTCAAGTTTGATGACTA 449
QY 308 CACACTGGTGTGCGAAATGCTGGGATCTTGAATCGAGACTAGATTTGATCTCTTG 367
DB 450 CTCACGTGGTGGTTCGAAAGATCCGACATCGAAGACGACCCGATCATCATCTCG 509
QY 368 AAGATGATGATTTCTGTGACACCGTATGTTTCTCTTCCGATCGTTGAGGACGCTTC 427
DB 510 AAGACCTCGGTCGGTCTCTTATGTTCTTCTGCGCTTCTTGAAGACCTCTTTC 569
QY 428 GAACCTGATAGGCTGGGATGATGACTTGTGCTGTTTGTGTCGAGAGTGTCT 487
DB 570 GTTTCATCTTTAAGCTGGGGAAGACGATGACGTGGGTTGTGTCGATTCGAGGTCA 629
QY 488 CGAAGGTTGATGATGATCTCCGAAGTATGTTGATCTTCATGCTGTGATCTCGT 547
DB 630 CCGAGGTGACCGGGTCCGAGTTTCTGCAAGTTGTATGTTTCAACCGCGAAGATCCGT 689
QY 548 TTGCACTTTTAAAGAGGATGAGATGCTGAGGACCATCTTGAACCTTTTCCGTTGT 607
DB 690 TCAAGTCTGTGAAGACGCGATGAGGTGTTAGGGTTATATGAACACCTTCAACCTGT 749
QY 608 TGGAGAGAAAGCTCCACGAGTATCGTGACAATTCGGTTGCTGACGTGGAGACGT 667
DB 750 TGGAGAGAAAGRCGCGCGGGAATGCTGATAGTTTCGGGTGCTGACGTGGAGACGT 809
QY 668 TTATACCTTAACGTTCAACAGGCGCTAATAGAAAGCTGAGGACCTCTGTCGACGCG 727
DB 810 TTATATTGACGTTGAACCTTGAACGAGTTCAATAGGGTGTAAAGTGTCTGTCGACGCGT 869
QY 728 GTTGTCCCTCGGTTTATGCTTAATAGAGATGTTGGAAATTCACGGAACAGATCGG 787
DB 870 GTTGTCCGCGGATGTTGCTTAATGACGAGGTTGGAAATGATTTGACATGATCTCG 929
QY 788 ATCCCATCAACAAGAAATGAACCAAAACCGTCCGCGGCAAAATGCCCTCGCTC 847
DB 930 ATGATATGATGTGAAGGATGAGTTTACCGTCCGCGGGAACAAATGCCCTTGAAGC 989
QY 848 TTTTGAATTTCCAAAGAAATTAACAAATTCGCTACTACTCAATCCCAAGGCCACCGGCC 907

DB 990 TTCTGAATTTCAAGAACTTCAAGTTTCAGAGACTACGTCCTCCGAAGACA----- 1043
QY 908 CCCGAGCCGCGCAGAAAGGATGAAGCGCTTATATGATGAACCTCAAGAGAGATTAA 967
DB 1044 -----AAACCAAGTCCGGAGTGAAGCTTCTGTAGAGATTTGAAGAAATTCCTCA 1097
QY 968 CTGTGAGCATTTTATGTTTGGATGCTTGTGATATTTGGAGTGGGCTTCCGCCGC 1027
DB 1098 CCGTGTATTACATCTACGCTGTGGACGCGCTTGTGGGCTACGTGGGTGTCTTCCGCC 1157
QY 1028 AGTGCCTGGCTTGGCTGAGCAGCTGTGATTCAGCCAGTGTCTTACACGAGGCTGACA 1087
DB 1158 GAGCTCTACTCTTCCGCCCTCAACTATTTGTCGGCCAGAGCTTCCGCCGGCTTAAAT 1217
QY 1088 TCAGCATGAGATTTTGGCGGTGATTAAGATGTTCTTCAATTAAGTCCGGGTGCTCCG 1147
DB 1218 TCAGCATGCAAGATTCGCGCTTGAATGAATGTGATACCGGAATTCGGATTCGTCG 1277
QY 1148 CGAAGAGCTGAGGATGTACGAAGCTTCATGCTCATTTTGGAAAAAGTTGGGATCG 1207
DB 1278 CGGACATGCGCAATGATTTTACGAAGCTTCTACTCTATCTTCAAAACGTCGGATTTG 1337
QY 1208 ACGGTATTAAGATGACGTTATTCACCTAATTTGAGATGTTGTGAAGACTATGAGGGA 1267
DB 1338 ACGCGTTTAAAGTTGACGTCATCAATATTTGAGATGTTGTGAGAAATATGGCGGGA 1397
QY 1268 GAGTGGATTTGGCAAAAGCATATTCAAAAGCAATGACCAATCAATTAATTAACATTTTA 1327
DB 1398 GAGTGAAGCTTGAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAGCTTAAAG 1457
QY 1328 AAGAAATGAGTCAATTCGCAATGATGAACATTTGAACACTTCAATGTTCTTGGACAG 1387
DB 1458 ACGGTAAACGCGCTTATCGGTACATGAGACACTGATATTTATTTTCTTGGAAACG 1517
QY 1388 AAGTATCTCTGTGCTGCTGTTGATGACTTTTGTGTCAGGACCCCTCTGTGATC 1447
DB 1518 AAGCATCTCTAGGTGCTGCTGATGATGACTTTTGTGTCAGGATCATCAGCGACA 1577
QY 1448 CAAGGCTACGTTTGGCTCCCAAGGATGTCACATGTTTATGTCGCAAGGACGACTGT 1507
DB 1578 TAAAGGACATATTTGGCTGCAAGGATGTCACATGTCACATGTCATCAACACTT 1637
QY 1508 GGATGGGCACTTCAATCCACCCCTGACTGGATGATTTTCAATCCACCCCTTGTGCG 1567
DB 1638 GGATGGGAATTTCAATCCACCCCTGACTGGATGATTTTCAATCCACCCCTTGTGCG 1697
QY 1568 CCTTCATGCTCTCTGAGGCAATCTGTGAGCCCATCTATGTTTACTGATCTCTG 1627
DB 1698 AGTTCATGCTCTCTGTCGATCTCCGCTGAGCCCATTTTACATCAGCATGTTGTG 1757
QY 1628 GAAGCATTAATTTGATCTTCTGAAAAAATCTGCTTCTCTATGATGATCTTCA 1687
DB 1758 GCGAGCAGATTTGATCTTCTGAAAAAATCTGCTTCTCTATGATGATGATG 1817
QY 1688 GTGAGTACTATGCTCCGCTGCTGATTTTGTGTTTGAAGACCTTGTGATTAAGAG 1747
DB 1818 GTGAGCAGATGCTACCTCCCACTCGTGAACGCTCTTTGAACACCTCTTCAATGAG 1877
QY 1748 AAAGTATGCTTAAGATTTGGAATCTCAACAGTTCACTGAGATGATGCTGATCAACT 1807
DB 1878 AAACCATGCTCAAGATTTGGAACCTTGAACACGATCACTGGAATTTTGGAGCATTC 1937
QY 1808 GCGAAGGAGAGATGATGCTGAGAGACGCGGCAACCAATGCTTTTACAAATCTCA 1867
DB 1938 GCGAAGGAGAGATGATGCTGAGAGACGCGGCAACCAATGCTTTTCCATGCTT 1997
QY 1868 AAGAGTGAATCAAAATTAACCCCAAAAGACATAGATGACACAGTGGAGAAACCTTA 1927
DB 1998 ACAGGTTAACCGCCCAACAAATTCCTTAAGAGCTTGAATGAACAGTGGAGAACCCCA 2057
QY 1928 TCTCTATTGAAGCGTTAAACCTTTTACCTTATCAAGGCAAAAACCTTATCC 1987


```
Db 958 ACAGATTGTAGTGGAAAGGATTCTG-----AGAAAGGATGGTGCTT 1002
Oy 939 TTTAGATGAACCAAAAGAGATTAAAGCTGTGGAGCATGTTTATGTTTGGCAGCTTT 998
Db 1003 TTTTAGGCACTTGAAGGAACAATTTAGAGCGTGGAGCGTGTATGTTGTGACCGCGT 1062
Oy 999 GTGTGGATATTTGGGGTGGCCTTCGCCGAGGTGCTGCTTGCCTGAGGACGCTGTGAT 1058
Db 1063 TTTGTGGATTTGGGGTGGGCTCAGACCCAGGTTCCGGGCAATGCCCAAGCTTAAGTTCT 1122
Oy 1059 TCAGCCAGTCTTTTCAACGAGGCTGCAGATGACATGAGATTTGGCGGTGATTAAGAT 1118
Db 1123 CACTCCGAAGCTGTCCAAATGAGCTAAATTTGACAAATGAAGATTAAAGCGTGAATAGAT 1182
Oy 1119 TTTTCTCTAATAGGTGGGGCTGCTCCCGGAGAAAGTGAGAGATGTAGAGAGACT 1178
Db 1183 CCTCAGTAAAGGAGTTGGACTGTGTCACACACCTGGCTCACTTTTGTATGACGGGCT 1242
Oy 1179 TCATGCTCATTTTGAAGAAAGTTGGATCGACGCTGTTAAGATTGACGTTATCCACTATT 1238
Db 1243 CCAGTCGCCGTTTGAATGTGGGGATTTAGAGGTGTTAAGTTGACGTTATACACTTGT 1302
Oy 1239 GGAGATGTTGTGAAGACTATGAGGAGAGAGTGGATTTGCCAAAGCATATTACAAAC 1298
Db 1303 CGAGATGCTATCCGAGGATACGGTGGCGGTGTGAGCTAGCCAAAGCTTATTACAAAGC 1362
Oy 1299 AATGACCAATATGAATAAATATTTTAAAGAAATGAGATTTGCAAGTATGGAACA 1358
Db 1363 GCTCACTGCTTGGTGTAAGCAATTTTCAAGCAATGGGCTCATTTGCAAGATGAGAGA 1422
Oy 1359 TTTTAACGACTTCATGTCCTTGGCACGGAACCTATCTCTCTTGGCTGTTGTGTATGA 1418
Db 1423 TTTGATATGACTCTTTTCTCTGTTACGGAACCAATAGCCCTTGGCGGCTGAGAGATGA 1482
Oy 1419 CTTTGTGTCACGAGACCCCTGTGTGATCAAAAGCTTGTGCTTCAAGATGTCA 1478
Db 1483 TTTTGTGTGACTGTATCCCTGTGAGATCCAATGTGACATTTGCTTCAAGGGGTGCA 1542
Oy 1479 CATTGGTTCATTTGGCCAAAGCAGCTGTGATGGGGAACCTTCAACCCGAGCTGAGCA 1538
Db 1543 CATTGGTCACTGTGCTTCAACAGCTGTGTGATGGGGAATTTTATTCACCGGATTTGGGA 1602
Oy 1539 TATGTTCCAATCCACCAACCTTGTGCGCTTCCATGCTGCTCTGAGCCATCTCTGG 1598
Db 1603 CATGTTCCAGTCCACACTCACCTTGTGCGGAATTTCCATGCGNGCTTAGGGCCATCTCTGG 1662
Oy 1599 TGGCCGATCTATGTTAGTGAATTTGTGGAAAGCATTAACCTTGTCTGAAAAAAT 1658
Db 1663 TGGACCAGTTTACGTTAGTGAATTTGTGGAAAGCACAACCTTCAAGTCTCAAGAGCTT 1722
Oy 1659 AGTGGTTCCTGATGATGATGCTTTCGAGTGAATGATGATGATGATGATGATGATGATG 1718
Db 1723 CCGTTTGGCTGATGGAGCATTTTGGCTGTGACACATGATGATGATGATGATGATGATG 1782
Oy 1719 TTTGTTTGAAGACCCCTTTCATATGAGAAATATGATGATGATGATGATGATGATGATG 1778
Db 1783 TTTGTTTGAAGACCCCTTTCATATGAGAAATATGATGATGATGATGATGATGATGATG 1842
Oy 1779 GTTCACTGAGTGAATTTGTGATTAACATGCAAGAGAGAGATGGTGTGAGACAGC 1838
Db 1843 AATATACAGGTGTTTGGGTCTATTTAATTTGCAAGAGAGTGGTGTGTCGCCGTACTG 1902
Oy 1839 CCGCAACCAATGCTTTTCAACAATATCAAAAGAGTGAATCCAAACATTAACCCAAAGA 1898
Db 1903 GAGAAACAAGAGTGGCTGCTTTCACAAACTGTGACATGCTTAAGCGAGTCTTCAAGA 1962
Oy 1899 CATAGATGGCAGAGTGAAGAAACCCATATCTATTTGAAGCGCTTAAACCTTTGCGCT 1958
Db 1963 CATAGATGGCAGAGTGAAGAAACCCATATCTATTTGAAGCGCTTAAACCTTTGCGCT 2022
Oy 1959 TTAAGCTTATCAAGCAAAAACTTATCTCTTCAAGCCCTTCAAGATCTTGACATAGC 2018
```

```
Db 2023 ATATTTGTTCAAGACACAAACTAAAGCTCATGAAGCATCAGAAATTTGAAGTTTC 2082
Oy 2019 TCTTGACCATTTGCAATTTGAGCTCATGCTGTTTACACAGTGAACCAATCATCAAC 2078
Db 2083 ACTTGAGCCATTTACTTTTGAACCTTATGACAGCTGTCTCAGATGATTTGCTGTCAAAA 2142
Oy 2079 TTTCTTACACTTTTGGCCCAATTTGGCTGTGATGACATGCTTAAACACTAGTAGAGCCATCA 2138
Db 2143 GTTAATTCATATTTGCTCCAAATTTGATTAATGTAACATGCTTAAACATGATGGTGGTCCATCA 2202
Oy 2139 ATCTGTGACATTTGACGATGACCTTAAGCTCAAGTCAAGTGTGCTCAAAAGGCTGTGTA 2198
Db 2203 GTCCATGTGAGTTTGAACAACACATGATGTGCTCAAAATTTGGGTTAGCGTTGGGGA 2262
Oy 2199 GATGCGAGTATTTGCATGCAAAAAACCAAGGCTGTGCTATGATTTGGGAGGATGTTGG 2258
Db 2263 GATGAAGGTGTTTGCATGAGCAAAACCAATGTTAGTTTCAAACTAGATGGGTTAGTTTAA 2322
Oy 2259 GTTCAAGTATGATCAAGCAAAATGTTGGTGTTCAGTGCATGCGCAATTTGATTTCTTC 2318
Db 2323 ATTTGATTAATG---AGGATTAATATGCTGAGAGTGCAGATGCTGCTGGCTTCAAA 2379
Oy 2319 ATCG 2322
Db 2380 ATTG 2383
```

RESULT 5

AAC89523

ID AAC89523 standard; DNA; 2498 BP.

XX AAC89523;

XX AC 07-MAR-2001 (first entry)

XX DE Soybean raffinose synthase coding sequence SEQ ID NO: 23.

XX KW Plant promoter; transgenic plant; desired property; ds.

XX OS Glycine max.

XX PN EP1048733-A2.

XX PD 02-NOV-2000.

XX PF 27-APR-2000; 2000EP-0108962.

XX PR 30-APR-1999; 99JP-0124527.

XX PR 01-SEP-1999; 99JP-0247211.

XX PA (SUMO) SUMITOMO CHEM CO LTD.

XX PI Ishige F, Watanabe E, Oeda K;

XX DR WPI; 2001-104537/12.

XX P-PSDB; AAB49400.

XX PT New soybean plant promoters useful for generating transgenic plants

XX PS with desired properties -

XX CC Example 2; Page 24-28; 36pp; English.

XX CC The present invention provides novel plant promoters which can be used in

XX CC the production of transgenic plants which express genes with desired

XX CC properties.

SQ Sequence 2498 BP; 651 A; 521 C; 673 G; 653 T; 0 other;

Oy Query Match 41.5%; Score 976.4; DB 22; Length 2498;

Best Local Similarity 66.0%; Pred No. 1.3e-303;

Matches 1507; Conservative 0; Mismatches 741; Indels 36; Gaps 5;

Oy 57 TGACATGTCGTCACCGTTTGAATGACGAGATCGGATTTCACTGTGAACGGTCAATTCGTT 116

118 TAATTTGCTTTTGTCCATAAACCCCTAGAGGATCAAAATTTCTCGCCAAACGGACCCCTTT 177
117 TCATGCGATTTCTCGAGAAACATTTGTGCTTCCCTCCGCTAC-----AC 164
178 TCTCAGGAAAGTTCCGAAACATAATAGTCACCCCTTCACCCATGACGCCAAGACTAG 237
165 TTGATAGACAAAGTCCCGGTTTCGGTTGGTCTTTGTGATTTGACACCGCTCGAACC 224
238 TAGAACAACAGAGAGACGACACGTCGTAGTTGCTTCTGTTGGCTTCCACGGACGAGCC 257
225 TGAATAGCCAGATGTTGTTTCAATTTGGGAAGCTGAAGATATTCGGTTATATGATTTT 284
298 CACAAGCCGACAGTGGCTTCCCTGGGGAAGCTCAGAGAAATTAATTCATGAGCATATTT 357
285 CAGTTTAAAGGTTTGGTGGCTACACATGGGTTTGGTGAAGTGGGAGGATCTTGATTC 344
358 CCGGTTTAAAGTTGGTGGACCACTCAGTGGGAGGATGACACGACACTGGAGCA 417
345 GGAGACTCAGATTTGTATCTTGAAGAGTCAG---ATTCTGTGACCGCTATTTTCT 401
418 CGAGACACAGATGATGCTTCTGCAAAAAGACACAGCTGGAGCCCTTTGTGTTGAT 477
402 TCTTCGATGCTTGAAGGACCGTTCCGACCTGATTCAGCTGGGAGATGATCTTTGT 461
478 TCTCCGATCTCCAAAGCCCTTCCGAGCCCTCCGCAACCCGTTTGGATATTACGT 537
462 CGATGTTTGTGCGAGAGTGTGTCGTAAGAAAGTTGTTGATGATCGTTGCGAAGTATGT 521
538 GGAAGCTTTGATGAGAGCGGAGTCGACAGCTGTCTGTGCTCCAGCTTCGGAGCTGCT 597
522 GTATCTTCATGCTGTGATGATCCGTTTGCATGTTTAAAGAGCGATGAATCTGTAG 581
598 ATAGCTCCAGTGGCATGCCGATACGTTGCTTAGAGAGCAACTAAATCGTTAG 657
582 GACCCATCTTGAACCTTTTCGTTTGGAGAGAAAGCTCCACCGATATCTGTGACAA 641
658 GATGCAATTTGGGAGCTTCAAGCTTCTCGAGGAAAAACCGCGCATGTCTATAACAA 717
642 ATTGCTGTGTGACAGTGGGACGCTTTTACCTAACGTTTCAACAGGACCTAATAGA 701
718 GTTGTGTGGTATCATGTGGAGCGCTTTTACTTGAAGGTGCATCCCTCAGAGTGTGGGA 777
702 AGGCGTGAAGCATCTGTCGACGCGGCTTCTCCCGGTTTATGCTCTAATGACGATGG 761
778 AGGGGTGAAGAGGTTGTTGGAGGAGGTGCCCTCCAGGGATGTCTAATCACAACGG 837
762 TTGGCAATCATGAGACGATTCGATCCCATCAC---CAAGAGAGATGAACCAAC 818
838 GTGGCAAGCATTTTGTACGACGAGAGACCCCATTAACGACCAAGAGGATGTGAAGCGAAC 897
819 CGTGGCGGAGCAAAATGCCCTGCTTTTGAATTTCAAGAGATTTACAAATTCG 878
898 CTCGCCAGGGAGCAAAATGCATGCAAGTTGTGAAGTTGGAGAAATTTACAAAGTTGAG 957
879 TGACTATGCTATCCCAAGGCCACCGGCCCGGACCGGACAGAGGAGGTAGAGCCGTT 938
958 ACAGTATTTAGTGAAGAGATTCG-----AGAAAGTATGGTCCCTT 1002
939 TATAGATGAACCAAGAGAGATTTAAGACTGTGAGACATGTTATGTTGGATGCTTT 998
1003 TGTTAGGAGCTTGAAGAAACAGTTTAGAGACGTGGAGCGAGTGTATGTGGACCGCT 1062
999 GTGTGATATTGGGAGTGGCTTCCGCGCAGAGTGCCTGGCTTTCGAGCAAGTGTGAT 1058
1063 TTGTGGGTATTTGGGGTTCAGACCAAGGTTCCGGGCAATGCCACAGGCTAAGTTGT 1122
1059 TCAGCAGATGCTTTACCAAGGGCTGCAATGACGATGAGAGATTTGGCGTGAATAGAT 1118
1123 CACTCGAAGCTGTCCAAATGAGACTAAATTTGACAAATGAAGATTTAGCGTGAATAGAT 1182
1119 TGTTCCTTAAGGTGGGCTGTCCCGCGGAGAGGCTGAGAGATGTACGAAGACT 1178

1183 CGTCAGTAAACGAGTTGGACTGTGTCACCAACCACTGGCTCACCTTTTGTAGCAGGGGCT 1242
1179 TCATGCTCATTTGGAAAAAAGTTGGGATTCGACGGTGAATGATGAGTTATCCACCTAT 1238
1243 CCACTCCGTTTGGATCTCGGGATTCGAGGTGTTAAGGTTAGAGTTATACACTTGT 1302
1239 GGAAGATGTTGTGAAAGACTATGAGAGAGAGTGAATTTGGCAAGGACATATTACAAG 1298
1303 CGAAGTGTCTTCGAGGAATAGGTGGCCGCTTTGAGCTAGCCAAAGGCTATTACAAGC 1362
1299 AATGACCAATCAATTAATTAACATTTTAAAGGAATGAGTATTTGCAATGAGTGA 1358
1363 GCTCAGCTGCTTCGTGAAGAGCATTTCAAGAGCAATGGGCTATTTGCGAGCATGAGCA 1422
1359 TTGTAAGCATTCATGTTCTTGGCAGGAGGATATCTCTGTTGCTGCTGCTGATGA 1418
1423 TTTTAATGACTTTCTTCTCTTGGTACCAAGGCAATAGCCCTTGGGCGCTAGAGATGA 1482
1419 CTTTGTGACAGCAGACCCCTCTGTGATCCAAACGCTAGCTTTTGGCTCCAAAGATGA 1478
1483 TTTTGTGCACTGATGCCCTGTGAGATCCAAATGACAGTATTTGGCTCCAAAGGTGTCA 1542
1479 CATGTTCAATTTGTCACACAGCTTTGTGATGGGAACTTATCCACCTGACTTGGGA 1538
1543 CATGGTCACTGTGCTTACAAACAGCTTGTGATGGGAAATTTTATTCAGCCGATTTGGGA 1602
1539 TATGTTCAATCCACCAACCCCTGTGCGGCTTCATAGCTGCTTCGAGCCATCTGTG 1598
1603 CATGTTCACTGCTCAGCTCACCCTTGTGCCAATTCATGCAAGCTCTTAGGGCCATCTGTG 1662
1599 TGGCCGATCTATGTTAGTATTTCTGTGGAAGATTAACCTTTGATCTTCTGAAAAACT 1658
1663 TGGACAGTTTACGTTAGTATGTTGTGGAAGACCAACTTCAGTTGCTCAAGAGCT 1722
1659 AGTGTCTCCGATGATGATCACTCTGAAAGTGAATGATGATGATGATGATGATGATG 1718
1723 CGCTTTGCTGATGGAGCAATTTTGTGTTGCTCAACCTATGCACTCCCAACAGAGATG 1782
1719 TTTGTTTGAAGACCTTTGATATGAGAAACTATGATTTGAATTTGAATTCACAA 1778
1783 TTTGTTTGAAGACCTTTGATATGAGAAAGCAATGCTCAAAATTTTGAATTCACAA 1842
1779 GTTCACTGAGATGATGCTGATTCACAGTCCAGAGAGAGATGCTGTGTCGACACG 1838
1843 ATATACAGGTTGTTGGTCTATTTAATTGCCAAGAGGAGGTTGTGTCGCTAATCTAG 1902
1839 CCGGACCAATGCTTTTCAATACATCAAAAGAGGATGATCCCAAACTAACCCAAA 1898
1903 GAGAAACAAGAGTGCCTCTGAATTTTACACAACTGTGACATGCTTAGCGATCTCABA 1962
1899 CATGATGAGCAAGTGAAGAAAAACCTATCTCTATTGAAGGCTTAAACCTTTGCGCT 1958
1963 CATTAATGAGAGCAATGGGAAAAAGCCCAATATGCAATAAAGGATGAATGTGTCTGT 2022
1959 TTAATCTTATCAAGCAAAAACTTATCTCTCAAGCCCTCTCAAGATCTTGATAC 2018
2023 ATATTTGTTCAAGGACCAAACTAAAGCTCAATGAAGGATGAGAAATTTGAAATTTT 2082
2019 TCTTGACCATTCGATTTGAGGCTCAATGCTGTTTACCAAGGACCAAACTATCCAAAC 2078
2083 ACTTGAGCCATTTACTTTTGAAGTATTTGACAGTGTCTCAAGTATTTGTCTGTCAAAAA 2142
2079 TTTCTACACTTTTGGCCCAATTTGGGCTGTGTAACATGCTTAACTAGTGAAGCATCA 2138
2143 GTTAAATTAATTTTGTCCCAATTTGATTAAGTAAATGCTTAAACATGAGTGTGCAATTA 2202
2139 ATCTGTGACATATGACGATGACCTAAGCTCAGTGAATTTGTTCAAAAGGTTGTGTA 2198
2203 GTCCATGAGGATTTGACAAACCATAGATGTGTCAAAATTTGGGGTTAGGGGTTGTGGA 2262
2199 GATGCGATATTGCAATGAAGAAAAACCAAGGCTGTCTATTGATGGGAGGATGTTGG 2258
2263 GATGAAGGTTTGTGATCAGAGAAACGTTAGTTGCAAACTAGATGGGTTAGTTGTAAA 2322

QY	1188	TTTGAAAAAATTGGGATTCACCGCTGTGAAGTATGGGCTATTCACACCTATGAGAACTT	1247
QY	1188	TTTGAAAAAATTGGGATTCACCGCTGTGAAGTATGGGCTATTCACACCTATGAGAACTT	1247
Db	1423	TTTGAAAAAATTGGGAAATTGATGAGTCAAAAGTTGATGTCATCCATTTGTTGGAGATAT	1482
QY	1248	GTTGGAAGACTATGGAGGAGAGTGTGTCGCAAAAGGCAATATTACAAAGCAATGACCA	1307
Db	1483	GCGAGAGACTATGGAGGAGAGTGTGTCGCAAAAGGCAATATTACAAAGCAATGACCA	1542
QY	1308	ATCAATTAATTAACATTTTAAAGAAATGAGCAATTTGCAAGTATGAAACATTTGAACA	1367
Db	1543	ATCAAGTGTAGGATTTTCAAAAGCAACGGTGTGATTTGTACATGTGAGCACTGCAACA	1602
QY	1368	CTTATGTTCTTCCTGGACGGAGACCTATCTCTGTGCTGCTGTGGTGAATGACTTTGTG	1427
Db	1603	TTTTCATGCTCTTGTGACTGAGACCAATTTGTCTGTGCTGCTGTGGGATGACTTTTGGCC	1662
QY	1428	CACGAGCCCTCTGTGATATCCAAACGATGTTTGGCTCAGCAAGATGTCACATGTTTCA	1487
Db	1663	AACGATACCGCTGTGAAATTAATAGTACATATTTGGCTCCAAAGGCTGTCAATATGTGCA	1722
QY	1488	TTTGCCAACAGCACTTTGGATGGGAACTTCACCCAGCTGACTGGGATATTTGCA	1547
Db	1723	TTTGCCCAACAATGTCTTATGATATGGGAAACTTATACACCTGACTGGGACATGTTCA	1782
QY	1548	ATCCACCACCCTTGTGCGCCTTCCATGCTGCTCTGAGCCATCTGTGTCGCGAT	1607
Db	1783	ATTCACACCCCTTGTGCTAAATTTATGCTGATCTCTGCGATTTCTGTGGGACCAAT	1842
QY	1608	CTATGTTAGTATCTTGTGGGAAAGCATTAATCTTGATCTTGCAAAAAAATAAGTCTGCC	1667
Db	1843	TTATGTTAGTATTTTGTGGCAAGCATACATCCCTTGCTCMAAAGGCTGCTTGCC	1902
QY	1668	TGATGATCGATCTTGGAGTAGTACTATGTGCACTCCGCACTGCGATTTGTTGTTGA	1727
Db	1903	TGATGTTGCACTCTTGTGCGAGTAGTACCATGCACTCTACTAAGATTTGCTATTTGT	1962
QY	1728	AGACCCTTGATATATGAGAAACTATGCTTAAGTTTGGATCTTCACAAAGTTACATGG	1787
Db	1963	AGATCTTGTGACATGTGCMAAAACATGCTCAAAATTTGAAACCTCAACAAGTACAAATGG	2022
QY	1788	AGTATTTGGTCATTTCAACTGCCCAAGSAGSAGATGGTCTGTGAGACACCCGCAACCA	1847
Db	2023	AGTCTTGGAGTCTTCAATTTGCCCAAGSAGSAGGTGAGCCGTGAGTCTCGAAAAATCT	2082
QY	1848	ATGCTTTTCACATACACAAAGAGTGCATCCAAACCTAACCCAAAAAGACATGAAATG	1907
Db	2083	ATGTTTCTCAGAGTATTTCAAAACCTATTTCTGTGCAGCAACATGCCAAAGATGTTGAATG	2142
QY	1908	GCAAGTGTGAGAAAAACCTTCTATTTAAGGCGCTTAACCTTTGGCGTTTACCTTA	1967
Db	2143	GGAACAGGACACAAAGCCATTTCCCATCAAGAGTGGAAATTTTGGCATGTACTTAC	2202
QY	1968	TCAGCCAAAAAATTAATCTCTCTTCCAAAGCCCTGCAGAATCTTGACATAGCTTTGACCC	2027
Db	2203	CAAGAAAAAAGTAACTCTCTCACAACTATCTGCACACATTTGAATATCACTGTATACC	2262
QY	2028	ATTGCAATTAAGGCTCTACCTGTTTCACAGAGCAAACTCATTCACAAATCTGTCTACA	2087
Db	2263	CTTGATTTAGAGATTTATTTAGTCTCTCCGATGACAAATTTTACCCTGGAGTGCATGCC	2322
QY	2088	CTTTTGCCCCAATTGGGCTGTGTGAACATGCTTAACACTAGTGGAGCCATTCAAATCTGTGA	2147
Db	2323	ATTTGACACCATATGAGTAACTAATTAACATGTCTCAACGCGGAGGGGAGTCAAGTCTTTGA	2382
QY	2148	C-----TATGAGCATGACCTAAGCTCACTCGAGATTTGGTGTCAAGGGGTGTGTGAGAT	2201
Db	2383	CATCAGTAGAGATATATGAGATTAAGTGTTCACGTTGTATTTAAAGGGGCGGAGAAAT	2442
QY	2202	GCGAGTATTGTCACGAAAAAACCAAGGCGCTGTGTAATGATGGGAGAGATGTTGGATT	2261
Db	2443	GATGCTTTATTTCACTAGAAAAAGCCAAACCGTGTAGAGTAAATGAGAAAGACATGAGATT	2502
QY	2262	CAGTATGATAGGACCAAAATGTGTGT	2288

Db	2503	TCGATGATGAGAGCATGATTAGCT	2529					
RESULT 7								
XX	ID	AAH27438						
XX		AAH27438	standard; DNA; 2343 BP.					
AC		AAH27438;						
DT		17-AUG-2001	(first entry)					
XX		Soybean polynucleotide; SEQ ID 2.						
DE		Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;						
XX		plant; soybean; ds.						
KW		Glycine max.						
OS		JP2001078783-A.						
PN		27-MAR-2001.						
XX		03-JUL-2000; 2000JP-0200571.						
PF		09-JUL-1999; 99JP-0196036.						
PR		(SUMO) SUNITOMO CHEM CO LTD.						
PA		MPI: 2001-313373/33.						
DR		P-PSDB; AAB98659.						
XX		Novel mutant protein of raffinose synthase is useful for reducing the						
PT		raffinose oligosaccharide content in a plant body -						
XX		Disclosure; Page 20-23; 30pp; Japanese.						
PS								
XX								
CC		The present invention relates to a mutant protein of raffinose synthase						
CC		in which at least one aromatic amino acid present at the position of						
CC		about 1-7 amino acids from the N-terminus is deleted or replaced. The						
CC		mutant protein can be used for reducing the raffinose oligosaccharide						
CC		content in a plant body. The present polynucleotide from soybean, was						
CC		used in the present invention.						
CC								
SQ		Sequence 2343 BP; 597 A; 493 C; 648 G; 605 T; 0 other:						
Query Match		41.1%; Score 966.8; DB 22; Length 2343;						
Best Local Similarity		66.0%; Pred. No. 1.6e-309;						
Matches 1508; Conservative		0; Mismatches 737; Indels 39; Gaps						
6;								
QY	57	TCACATGTCGTCACCGTTTGCATTCGACCGATTCGAGATTTCACGTGCAACGGTCATTCGTT	116					
Db	57	TAAATTTGCCCTTTGTCATTAACCTTAGAAGATCAATTTCTCTCGCAACGGCCACCTTT	116					
QY	117	TCTGTCGATGTTCTCGTAGAACATTTGCTTCTCTCTCCGTAC-----AC	164					
Db	117	TCTCAAGCAAGTTCCCGCAAAACATATATAGTACCCCTTACCCATTCGACGCCCAAGATAG	176					
QY	165	TTTCGATAGACAAGTCCCGGTTGCGTTGGTTGCTTTGTTGGATTTCGACGGTCGGAAC	224					
Db	177	TAAAGCAACAAGCGAGGACGACGAGCGTGTAGGTGCTTCGTGGCTTCCACGCGGACGAGCC	236					
QY	225	TCATGACCGCATATGTTGTTTCGATTGGGAAGCTGGAAGATATTGCTTATGAGATTTT	284					
Db	237	CAGAAAGCCGACACGTCGTGCTTCCCTGGGAAAGTCCAGAGGAATAAATTCATGAGCATATT	296					
QY	285	CAGGTTTAAAGTTTGGTGGACTACACACTGGGTTGTCGAAATGGTGGGATCTTGAATC	344					
Db	297	CCGGTTTAAAGGTGGTGGACACATCAGCTGGGTGGTAGCAACGACACGAACTGGACGA	356					
QY	345	GGAAGCTACGATTTTGATTCCTTGAAGAAGTCAG--ATTCTGGTCGACCGTATGTTTCTT	401					
Db	357	CGAGACACAGATGATGCTTCTTCGACAAAAACGACGCTCGACGCCCTTGTGTTGAT	416					

Oy	402	TCCTCCGATCGTTBAGGAGACCGTTCCGAACTTCGATTAC6CTGGGATGATGACTTTGT	461
Db	417	TCCTCCGATCGTTCCAAACCCCTGTTTCGACCTCCCTCGCAACCCGGTGTGGATTACGT	476
Oy	462	CGATGTTTGTGTGAGAGTGTTCGTGTCGAAAGTGTGTGATGCGTTCCGAATATCTT	521
Db	477	GGAGCTTTGCATGGAAGACCGCTTCGACAGCTGCTCTGCGCTCCACCTTCGGAGCTGTT	536
Oy	522	GTAATCTTCATGCTGATGATGATCCGTTTGCACTGTTTAAAGAGCGATGAAGATGTAG	581
Db	537	ATACGCTCCAGCTTGGCCATATCCCGTATGATGTTGCTTAAAGAAACACGTAAGTCTTAG	596
Oy	582	GACCATCTTGGAACTTTTGCCCTTGTGTGAGAGGAAGACTCCACAGATCTGGGACAA	641
Db	597	GATCATTGTGGGACTTTCAGAGCTTCTCGAGGAGAAACCGGCCACGATGATGACAA	656
Oy	642	ATTGCGTGTGTGCACGTCGAGACCGCTTTTACCTAACGCTTACACAGGGCGTAATGA	701
Db	657	GTTTGTGTTGGTGTACATGGGAGCGGTTTACTTGAAGTGTGATCCCTCAGGTGTGTGGA	716
Oy	702	AGCGCTGAGGCATCTGCTCGACGGCGGTTGTCTCCCGGTTTAACTTAAATGACATGG	761
Db	717	AGGGGTAAAGGGTGTGTGAGGAGGGGTGCCCTCCAGGGATGTCTTAATCGACAGCG	776
Oy	762	TTTGCAATATCTCGGACAGCATTCGGATGCCATCAC---CAAGAAAGGATGAACCAAC	818
Db	777	GTCGCAACGCAATTTGTGTACACGACGAGACCCCATACGACACCAAGAGGATATGAACGAC	836
Oy	819	CGTGGCGGCGAGCAAAATGCCCTGTCCGCTCTTTTGAATTCGAAGAAATTCAAATTCG	878
Db	837	CTCCGACAGGGAGCAAAATGCCATGACAGTGTGGAGTTGGAGGAAATTCACAGTTAG	896
Oy	879	TGACTACGCTAAATCCCAAGGCCACCGGCCCCGAGCGGCCACGAAGAGGATGAAGCCTT	938
Db	887	ACAGTATTTGTAGTGAAGGATTTCTG-----AGAAAGGATGTGGTGCTT	941
Oy	939	TATAGATGAACCTCAAGAGAGATTTTAAAGACTGTGAGCATGTTTATGTTGGCATGCTT	998
Db	942	TGTTAAGGACTTGAAGGACATGTTTAGAGCGTGGAGCGAGCATGTATGT---GCAGCGCT	998
Oy	999	GTTGTGATATTTGGGTGGCTTTCGCCCGCAGTGCTGCTTGCCTGAAGCAGCTGTGAT	1058
Db	999	TTTGTGGGTATTTGGGTGGGTGTAGACCCCAAGATTCGGGCAATGGCCCAAGGCTTAAGCTGT	1058
Oy	1059	TCACACCATGCTTTCACACGAGCTGCGAGATGACAGTGAAGATTTGGCGGTGATTAAGT	1118
Db	1059	CACGCCCAAGCTGTCCAAATGACTAAATTTGACAAATGAAGGATTTAGCGGTGATTAAGT	1118
Oy	1119	TGTTCTTCATTAAGTTCGGGCTGTCCCGCGGAGAAAGCTGAGAGATGTACGAAGACT	1178
Db	1119	CGTCAGTAAGCAAGATTTGAGCTGGTGGCACACACCTCGCTCACTTTGTATGAGAGGCT	1178
Oy	1179	TCATGCTCATTTGGAAAAATTGGGATTCACAGGCTGTAAAGTTAGTACGTTATCCACTTT	1238
Db	1179	CCACTCCCGTTTGGAACTTCGCGGTATGTACGGGTGTTAAGTTGACGTTATACACTGCT	1238
Oy	1239	GGAGATGTTGTGTGAAGACTATTGAGGAGAGTGTATTTGGCAAAAGCATATTACAAGC	1298
Db	1239	CGAGATGCTATTCGAGGAATAACGTGGCGGCTGTGAGTACGCCAAAGCTTATTACAAGC	1298
Oy	1299	AATGACCAATCATTAATTAACATTTTAAAGAAATGGAATTCATTGCAAGTATGGAACA	1358
Db	1299	GCTACACTGCTTCGGTGAAGAGCATTTCAAAGCAATGGGGCTATTGCGAGCATGTGACAA	1358
Oy	1359	TTGTGATGAGACTTCATGTTCCCTTGGACACGGAATATCTCTTGGTCTGTTGGTGTGAT	1418
Db	1359	TTGTGAATGACTTCTTCTCTTGGTACCGAAGCCATTAAGCCCTTGGGGCGGTAGAGATGA	1418
Oy	1419	CTTTTGTGTGACGAGACCCCTCTGTGTATCCAAAGCATGTTTGGCTTCAGAGATGTCA	1478
Db	1419	TTTTTGTGTGCACTGATCCCTCTGTGAAGATCCAAATGGCACAGTATTTGGCTTCAGAGGCTGTCA	1478

QY	1479	CAGGTTTCATTGCGCAACAGACAGCGTTGGGAGTGGGAACTTCATCCACCCCTGACACTG	1538
Db	1479	CATGGTCGACGTGCTGCTACCAACAGCGTTGGAGTGGGAAATTTTAACTCAGCGGATG	1538
QY	1539	TATGTTCGAATCCACCCACCCCTGTGTCCGCTTCCTCCATGCTGGCTTCGAGCCATCTCG	1598
Db	1539	CATGTCACAGTCCACCTCACCCCTGTGTCCCAATTCATGACGCTCTAGGGCCATCTCG	1598
QY	1599	TGGCCCGATCTAGTGGTAGTATTCGTGTGGGAAAGCATTAATCTTTGATCTTGTGA	1658
Db	1599	TGGACCAATTTAGCTGATAGTATGTTGTGGAAAGCAACATTCAGATGTGCTCAAGAG	1658
QY	1659	AGTGTCTTCATGATGATTCGATCCTTGGAAAGTAGTACTATGACATCCGACTCCGAG	1718
Db	1659	CGCTTTCGCTGATGGGAGCAATTTTGGGTGTCCACACTATGACATCCGACACAGAG	1718
QY	1719	TTTGTGTAAGAGCCCTTGGATATAGGAGAACTATGCTTAGATTTGGATATCTCA	1778
Db	1719	TTTGTGTAAGAGCCCTTGGATATAGGAGAACTATGCTCAAAATTTGGATATCTCA	1778
QY	1779	GTTTACTGAGAGTATGTTGGTTCATTTCAACTGCCACAGAGAGGATGTTGCTGAG	1838
Db	1779	ATATACAGAGTGTTTTGGGTCTATTATATTTGCCAAGAGGTGGGTGTCTCCCTTA	1838
QY	1839	CCGCAACCAATGCTTTTACATATCTCAAAAGAGTGCATCCAAACTAACCCAA	1898
Db	1839	GAGAAACAAGAGTCCCTCTAATTTTACAAACTGACATGCTTACGAGAGCTCA	1898
QY	1899	CATGATATGCGACAGTGGAGAAACCCCTATCTCTATTTAGAAAGCGTTAAACCT	1958
Db	1899	CATGATATGCGACATTTGGGAAAGCCCAATATGCAATAAAAGGATGAATGTTG	1958
QY	1959	TTTACCTCATCAACCCAAAAACCTTATCTCTCCCAAGCCCTCTCAAGATCTTG	2018
Db	1959	ATATTTTGTTCAGAGCACCAAACTAAAGCTCATGAAGGATATAGAAATTTGA	2018
QY	2019	TCTTGACCAATTCGAATTCGAGCTCATCTGTTTCCACAGTAGACCAACATCA	2078
Db	2019	ACTTGAAGCAATTTACTTGTGAGCTATTTGACAGTGTCTCCAGTAGTGTGTCA	2078
QY	2079	TTTCTCTACACTTTGGCCCAATTTGGGCTGGTGAACATGCTTAACACTAGTGA	2138
Db	2079	GTTTAAATCAATTTCTCCATTTGGAATAGGAACATGCTTAACACTGGTGTGC	2138
QY	2139	ATCTGTGACATGATGATGATACCTTAAGCTCAGTCGATGTGTTGTCAAAGG	2198
Db	2139	GTTCCATGAGATTTGACACACCACTAATGTTGTTCAAAATTTGGGTTAGGG	2198
QY	2199	GATCGAGATATTTGATCGAAGAAACCAAGCGGCTGTCTTATTTGATGGGAG	2258
Db	2199	GATGAAGGTGTTTTCATCAAGAAACCAAGTATGTTGCCAATCTAAGTGGG	2258
QY	2259	GTTCAAGTATGATCAGGACCAAAATGTTGTTTCAAGTGCATGCGCAATGAT	2318
Db	2259	ATTTCATTAATG---AGGATAAATGCTGAGAGTGCMAAGTTCCCTGAGT	2315
QY	2319	ATCG 2322	
Db	2319	ATTG 2319	
RESULT 8			
AAZ10002			
AAZ10002 standard; cdna to mRNA; 2497 BP.			
XX	AAZ10002:		
XX			
XX	26-OCT-1999 (first entry)		
XX			
DE	Nucleic acid encoding a raffinose synthase protein.		
XX			
XX	Raffinose synthase; plant; sucrose; raffinose; ss.		
XX			

OS Glycine max. Location/Qualifiers
FH Key 62..2406
FT CDS /tag= a
FT /product= "raffinose synthase"
XX JP11215984-A.
XX PD 10-AUG-1999.
XX 12-DEC-1997; 97JP-0342899.
XX 28-NOV-1997; 97JP-0329006.
XX PR 18-DEC-1996; 96JP-0338673.
XX (SUMO) SUMITOMO CHEM CO LTD.
XX WPI; 1999-511112/43.
XX DR P-PSDB; AAY30143.
XX
XX New raffinose synthase gene - is prepared from a plant material
XX
XX Claim 9, Page 27-30; 40pp; Japanese.
XX
XX The present sequence encodes a raffinose synthase protein. The
XX sequence is isolated from plant material. The protein forms raffinose
XX by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C
XX of D-glucose residue in sucrose molecules.
XX
SQ Sequence 2497 BP; 650 A; 521 C; 673 G; 653 T; 0 other:

Query Match 41.1%; Score 966; DB 20; Length 2497;
Best Local Similarity 66.0%; Pred. No. 3e-300;
Matches 1507; Conservative 0; Mismatches 740; Indels 37; Gaps 6;

OY 57 TCACATGTCGTCACGCTTGGCAATGACGAGATCGATTTCACCTGTCAGCGTCATTGCT 116
DB 118 TATTTGCTTGTTCATTAACCTTAGAAGATCAATTTCTCCGCCACCGCCCTTT 177
OY 117 TCTGTCGAGATGTCGCAACATGTTGCTCTCTCTCCGTAC-----AC 164
DB 178 TCTCAGGAGATGTCGCAAAACATTAATAGTCAACCTTCACCCATGACGCCAAGGTAG 237
OY 165 TTCGATGACAAGTCCCGGTTTCGGTTCCTTGTGTTGATTCGACGCGTCGGAACC 224
DB 238 TAGAACAACGAGGAGACGAGCGTGTAGTTCCTGTCGGCTTCACGCCGACGACGCC 297
OY 225 TCGATGCCACATGTTGTTTCGATTGGGAAGCTGAAGATATTCGTTATGATATTT 284
DB 298 CAGAAGCCGACAGTGGCTTCCTCGGGGAAGTCAGAGGAATMAAATTGATGACATATT 357
OY 285 CAGGTAAAGTGTGGTGTGACTACACATGGGTGGTGAATGAGGGGATTTGAATC 344
DB 358 CCGGTTAAGGTGTGGTGAACACTCAGTGGGTGGTAGAAGGACGACGACTGGAGCA 417
OY 345 GGAGACTGCAATGTCATCTTGAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 401
DB 418 CGAGACAGATGATCTCTCTGCAAAAAACGACAGCTCGGACGCCCTTTGTTGAT 477
OY 402 TCTTCGATGTTGAGGAGCGGTCGGAACCTCGATTCAAGCTGGGGATGATGATTTGT 461
DB 478 TCTCCGATCTTCAAGCTCGTTCCGACCTCTCCCAACCGGTTTGGATGATTACGT 537
OY 462 CGATGTTTGTGTCGAGAGTGTGTCGGAAGTGTGATGATGATGATGATGATGAT 521
DB 538 GGAAGCTTTCATGAGAGGCGGGTGCACGCTGTCTGCTCCAGCTTCGGGAGCTGCTT 597
OY 522 GATCTTCAATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 581
DB 598 ATACGTCACAGTGTGGCATGACCCGATCATGTTGCTTAGAGAGCAACTAAATCGTTAG 657
OY 582 GACCCATCTTGAACCTTTGCTGTTGAGAGGAAAGACTCCACAGATATCTGGACAA 641

DB 658 GATGATTTGGGAGACGTTAAAGCTTCGAGAGAAACCGCGCGATATATAGACAA 717
OY 642 ATTGGTGTGTCAGCGAGGACCGGTTTACCTAAGCTTCAACAGAGGCGATAGAGA 701
DB 718 GTTTGGTGTGTACATGGGACCGGTTTACTTGAAGTGTGATCCCTCAGGTGTGGGA 777
OY 702 AGGCGTGAAGCATCTGTCGAGGCGGTTTCCTCCGCTTTAGTCTTAATGACAGATGG 761
DB 778 AGGGGTGAAGAGGTTGTGAGAGGAGGTGCTCCAGGAGATGCTCTAATGACGACGG 837
OY 762 TTGGCAATCATTCGACAGATTCGATCCCATCAC---CAAGAAGGATGAGCAAC 818
DB 838 GTGGCAAGCATTTGTACAGACAGAGACCCCATAGACGACGAGGATGAGAGGAC 897
OY 819 CGTCGCCGCGAGCAATGCCCTGCTGTTTGAATTTCAAGAGATTTACAAATTCG 878
DB 898 CTCGCGAGGAGGACCAATGCCATGACGTTGGTGAAGTTGAGAGAAATTTACAGTTAG 957
OY 879 TGACTAGTCATCCCAAGGCCACGGCCCCGAGCCGCCAGAGGGGATGAGGCGTT 938
DB 958 ACAGTATTTAGTGAAGGATCTG-----AGAAAGGATGGGTGCTT 1002
OY 939 TATAGATGAACCTCAAGAGAGATTTAAGCTGTGAGCATGTTTATGTTGGCATGCTT 998
DB 1003 TGTTAGGACTTGAAGAACAGTTTAGAGCGTGGAGCAGGTGTATGTGACCGCGCT 1062
OY 999 GTGTGATATTGCGGGGTGCGCTTCGCCGAGGTGCTGCTGCTGAGGACGCTGTGAT 1058
DB 1063 TTTGGGTATTGGGTGGGTGAGGTCAGACCAAGTTCCGGGATGCCAGGCTTAAGTTGT 1122
OY 1059 TCAGCCAGTCTTTCACACAGGCTGCAATGACGATGAGAGATTTGGCGGTGATAGAT 1118
DB 1123 CACTCCGAGGCTTCCCAATGAGTAAATTAATGAAGATTTAGCGGTGATAGAT 1182
OY 1119 TGTTCATTAAGTGGGTGTGTCGCCGCGAAGGCTGAGAGATGATGACAGGACT 1178
DB 1183 CGTCAGATGAAGGATGTGACTGTGTCACACACTGCTCACCCTTTGTGACAGGGGCT 1242
OY 1179 TCATGCTCATTTGGAAGAAAGTTGGATTCGACGATGATTAAGATTTACCACTATT 1238
DB 1243 CCATCCCGCTTGGAAATTCGCGGTAATTGACGTTAAGGTTGAGCTTATACACTTGT 1302
OY 1239 GGAATGTTGTGTAAGACTATGACGAGGAGTGTGATTTGGCAAGGATTTACAAAGC 1298
DB 1303 CGAGATGCTATCCGAGGAATACGTCGCTGTTGAGCTAGCCAAAGCTTATACAAAGC 1362
OY 1299 AATGACCAATCAATTAATTAATTAAGGAATGGAATGGAATGGAATGGAATGGA 1358
DB 1363 GCTCAGCTGCTGGTGAAGAGCAATTTCAAGGCAATGGGTATTTGCGGATGAGCA 1422
OY 1359 TTGTAAGACTTCATGTTCTTGGACAGGAGCTATCTCTTGTGTCGTTGGTGTATGA 1418
DB 1423 TTGTATGACTTCTTCTCTCTGTTACGGAAGCCATAGCCCTTGGGGGCTAGAGATGA 1482
OY 1419 CTTTGGTGCAGGACCCCTCTGTGATCAAAAGTACGTTTGGCTCCCAAGAGTCA 1478
DB 1483 TTTTGGTGCAGTATCCCTCTGAGATCCAAATGACCTATTTGGCTCCCAAGGCTCA 1542
OY 1479 CATGTTGATTTGGCAAGCAGCTTGTGATGGGAATTCATCCACTGACTGAGGA 1538
DB 1543 CATGTTGATTTGGCAAGCAGCTTGTGATGGGAATTTTATTTATTCAGCGGATTTGGGA 1602
OY 1539 TATGTTCAATCCACCCACCTTGTGCGGCTTCCATGCTGCTTCGAGCAATCTCTGG 1598
DB 1603 CATGTTCAATCCACCTGATCACTGCTGCGGATTTCCATGTC-GCCTTAAGGCCATCTCTGG 1661
OY 1599 TGGCCGATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1658
DB 1662 TGACACAGTTTACGTTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1721
OY 1659 AGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1718

Db 1722 CGCTTGCCTGATGGAGACATTTGCGTTGTCACACATATGACATGCCACACAGACTG 1781
 Oy 1719 TTTGTTGAGAACCTTTGCATATATGAGAGAACTATGCTTAAGATTGGATCTCAACA 1778
 Db 1782 TTTGTTGAGAACCTTTGCATATATGAGAGAACTATGCTTAAGATTGGATCTCAACA 1841
 Oy 1779 GTTCAGTGAAGTATGATGTCATCTCACTGAGAGAGAGAGATGCTGTCAGACACG 1838
 Db 1842 ATATACAGGTGTTGGTGTCTTAAATTAATGAGAGAGAGAGAGAGAGAGAGAGAG 1901
 Oy 1839 CCGGACCAACCTCTTTTACATATCTCAAAAGAGATGATGATCCAAACCAACCAAGA 1898
 Db 1902 GAGAAACAGAGATGCTGCTTGAATTTTTCACAAAGCTGATGAGCAGATCTCAAGA 1961
 Oy 1899 CATGAATGAGCAGTGGAGAAACCTATCTCTATTTGAGAGAGAGAGAGAGAGAGAG 1958
 Db 1962 CATTAATGAGCAGTGGAGAAACCTATCTCTATTTGAGAGAGAGAGAGAGAGAGAG 2021
 Oy 1959 TTACCTCTATCAAGCCAAACCTATCTCTCAAGCCCTCTCAAGATCTTGACATAGC 2018
 Db 2022 ATATTTGTTCAAGGACCAACCACTAAGCTCATGAGAGCAGATGAGAAATTTGAGATT 2081
 Oy 2019 TCTTACCATTTGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATG 2078
 Db 2082 ACTTGAAGCATTACTTTGAGATTCGATGATGATGATGATGATGATGATGATGATGATG 2141
 Oy 2079 TTTCTTACCATTTGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATG 2138
 Db 2142 GTTATATGATTTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 2201
 Oy 2139 ATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2198
 Db 2202 GTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2261
 Oy 2199 GATGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2258
 Db 2262 GATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2321
 Oy 2259 GTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2318
 Db 2322 ATTTGATTTATG---AGATTAATGCTGATGATGATGATGATGATGATGATGATGATGAT 2378
 Oy 2319 ATCG 2322
 Db 2379 ATTG 2382
 RESULT 9
 AA40800 standard; cDNA to mRNA; 2746 BP.
 ID AA40800;
 AC AA40800;
 XX
 DT 23-SEP-1998 (first entry)
 XX
 DE Broad bean raffinose synthetase coding sequence.
 XX
 KM Raffinose synthetase; metabolism modification; food additive;
 KW gastrointestinal flora; broad bean; ss.
 XX
 OS Vicia faba.
 OS
 FH Key Location/Qualifiers
 FT CDS 101..2500
 FT /tag= a
 FT /product= raffinose synthetase
 FT
 PN EP849359-A2.
 XX
 PD 24-JUN-1998.
 XX
 PF 18-DEC-1997; 97EP-0122417.
 XX

PR 18-DEC-1996; 96JP-0338673.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 XX
 PI Oeda K, Wantanabe E.
 DR WPI: 1998-324670/29.
 XX P-PSDB: AAW57886.
 PT New nucleic acid molecule encoding plant raffinose synthetase -
 PT capable of producing raffinose, used as food additives with
 PT beneficial effects on gastrointestinal flora
 PS
 XX
 CC
 CC This sequence encodes the broad bean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.
 XX
 SQ Sequence 2746 BP; 780 A; 552 C; 621 G; 793 T; 0 other:
 Query Match 40.7%; Score 956.6; DB 19; Length 2746;
 Best Local Similarity 66.6%; Pred. No. 3.5e-297;
 Matches 1421; Conservative 0; Mismatches 694; Indels 18; Gaps 3:
 Oy 193 GCTTCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 252
 Db 359 GGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 418
 Oy 253 AAGCTGAAGATATTCGTTTATGATATTTTCAAGTTTGAAGTTTGAAGTTTGAAGTTTGA 312
 Db 419 AACTTAAAGATATTCGTTTATGATATTTTCAAGTTTGAAGTTTGAAGTTTGAAGTTTGA 478
 Oy 312 TGGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 372
 Db 479 TGGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 538
 Oy 373 TCAGATTC---TGGTGCAGCTGATGTTTCTTCTTCCGATGTTGAGGAGGACCTTCCGA 429
 Db 539 AAGCACTCCCTGAGAGACCTATGCTTACCTCCGATGTTGAGGAGGACCTTCCGA 598
 Oy 430 ACCCTGATTCAGCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 489
 Db 599 ACCCTGATTCAGCTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 658
 Oy 490 AAGGTTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 549
 Db 659 CATGTCACCGGCTCAAGCTTCAAGAGATGTTTCAATCCTCAGTACGACCATATAC 718
 Oy 550 GCATTTGTTAAGAGGCGATGAGATGCTGAGAGACCATCTTGAACCTTTTTCGTTTGTG 609
 Db 719 AGTATACGTTAAGAGGCGATGAGATGCTGAGAGACCATCTTGAACCATCTTGTG 778
 Oy 610 GAGGAGAGACTCCACAGAGATGCTGAGAGCAATTCGTTGTCAGCTGAGAGAGCTTT 669
 Db 779 GAGGAGAGAGACTCCACAGAGATGCTGAGAGCAATTCGTTGTCAGCTGAGAGAGCTTT 838
 Oy 670 TACCTAAGGCTTCAACAGGAGGCAATTAAGAGGCTGAGAGCTTCTGAGGAGGCT 729
 Db 839 TACTTGAAGGCTTCAACAGGAGGCAATTAAGAGGCTGAGAGCTTCTGAGGAGGCT 898
 Oy 730 TGTCTCCCGGTTTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 789
 Db 899 TGTCTCCCGGTTTGAATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 958
 Oy 790 CCATTCACCAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 849

Db 959 CATGAGATGATTCAGAGATGTAACCGAACCTCAGCCGGGGAACAAATGCCATCGACACTT 1018
 QY 850 TTGAATTTCCAGAGAAATTCAGAAATCCGCTGACTACGTCAATCCCAAGGCCACCGGCC 909
 Db 1019 GTAAATAGCAGAGAAATCTTAAGTTAGAAATATGAAATCTCTGA-----A 1066
 QY 910 CGAGCCGCCCAAGAGGATGGAAGCGCTTTATAGATGAACCTCAAGAGAGATTTAAGACT 969
 Db 1067 AATGAGAGGAAAGAAAGTTGGTGGTTTGTAGAGGATTTAAGAGAAAGTTTGGAGCT 1126
 QY 970 GTGAGACATGTTATGTTGGCATCTTGTGTGATATTGGGTTGGCTTGCCTGCGCAG 1029
 Db 1127 GTGAGAGAGTATTATGTTTGGCATCGCTTGTGGTATTGGGGCGGGTATAGCCTGGA 1186
 QY 1030 GTGCTGGCTTCCCTGAGGACGCTGTATTCAGCAGTCTTTCACCAAGGCGCTGCAGATG 1089
 Db 1187 GTGCTGGGATGCCAAAGCTAGGGTGTGTTCGAAAGGTCTCAGAGGGGTTGAAGATG 1246
 QY 1090 ACGATGAGAGATTTGGCGGTGATTAAGATTTCTTCATTAAGTGGGCTGGTCCCGCG 1149
 Db 1247 ACGATGAGAGATTTGGCGGTGATTAAGATTTGTGAGAAAGGTGGGCTAGTGGCGCA 1306
 QY 1150 GAGAAGCTGAGAGATGATGACAAAGCTTCATGCTCATTTTGGAAAAATTTGGATGCAC 1209
 Db 1307 GATTTTGCACATGAGATGTTTATGAGGCTTCACTCTCATTTTGGATCGCGGGAATTCAC 1366
 QY 1210 GGTGTTAAGATTGACGTTATTCACCTATTGAGATGTTGTGAAAGACTATGAGAGGAGA 1269
 Db 1367 GGTGTTAAGATTGATGTTATTCATCTGCTTGTAGTTACTATCAGAGAGAAATATGTGAGCA 1426
 QY 1270 GTGATTTGGCAAGGCAATATACAAAGCAATGACCAATCAATTAATTAACATTTTAA 1329
 Db 1427 GTTGAGCTAGCAGAGACTTATTACAAAGCAATACCTCAGTGAAGAAATTTTCAAA 1486
 QY 1330 GGAATGAGATCTTCAAGATGGAACATTTGTAACGACTTATGCTTGGCAGCGAA 1389
 Db 1487 GGCATGAGTATGATGCTTCAAGATGGAACATTTGTAACGACTTATGCTTGGCAGCGAA 1546
 QY 1390 GCTATCTCTGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1449
 Db 1547 GGCATATCCCTCGCGCGCGCGCGCGAGATGATTTTGGTCTGATTCATCTGATTCGA 1606
 QY 1450 AAGGTTAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1509
 Db 1607 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1666
 QY 1510 AATGGAATCTTCAATCCACCTGATGATGATGATGATGATGATGATGATGATGATGATG 1569
 Db 1667 AATGGAATCTTCAATCCACCTGATGATGATGATGATGATGATGATGATGATGATGATG 1726
 QY 1570 TTTCATGCTGCTGAGAGCCATCTCTGTGGCCGAGTATGATGATGATGATGATGATGATG 1629
 Db 1727 TTTCATGCTGCTGAGAGCCATCTCTGTGGCCGAGTATGATGATGATGATGATGATGATG 1786
 QY 1630 AAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1689
 Db 1787 AATCAAAATTTCAAGTGTCTCAAAATCTCTGTGGCCGAGTGTCTATCTTGGCTGT 1846
 QY 1690 GAGTACTATGACCTCCGAGCTGCGATGTTGTTTGAAGACCTTTGGATGATGATGATGATG 1749
 Db 1847 CAACATTTACGACCTCCCTCAACAGATGATGATGATGATGATGATGATGATGATGATGATG 1906
 QY 1750 ACTATGCTTAAGATTTGAATCTCAACAGTTCAGTGAAGTATGATGATGATGATGATGATG 1809
 Db 1907 ACAATGCTGAATAATTTGAATCTCAACAAATATACAGGTGTTTGGCTTCTTCAACTGC 1966
 QY 1810 CAGAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1869
 Db 1967 CAGAGTGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2026
 QY 1870 CGAGTACATCCAAAGTATACCAAGATGATGATGATGATGATGATGATGATGATGATGATG 1929
 Db 2027 GCGTGACATGTTATGCAAGTCCGAAAGCATTTGAATGTTGCAATGGCAAACTCCAAATG 2086

QY 1930 TCTATTTGAAGCGGTTAAACCTTTGCGCTTATACCTATACAAAGCCAAAAACTTATCTTC 1989
 Db 2087 AACACCAAGAGTGTGATTTTGTCTGTATTTTTCAGAGAGAAAGAAATGAGAGCTC 2146
 QY 1990 TCCAAAGCCCTCAAGATGTTGACATAGCTCTTGACCCATTTGGAATTTGAGCTACACT 2049
 Db 2147 ATGAGCTGTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2206
 QY 2050 GTTTCACAGTACCAACCTCAACCTTCTACACTTTTGCCTCAATTTGGGCTGCTG 2109
 Db 2207 GTGCTCCAGTGAAGATGTTTTCGAAAGGTTATACAGTTTGCACGATTTGGGTTAGTG 2266
 QY 2110 AACATGCTTAACACTAGTGGAGATCCATCTGTGAGATGATGATGATGATGATGATGATG 2169
 Db 2267 AACATGCTTAACACTAGTGGAGATCCATCTGTGAGATGATGATGATGATGATGATGATG 2226
 QY 2170 GTGAGATGTTGTGCAAGGTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2229
 Db 2327 GTCAAGATTTGGGTTGAGAGATGCGGGGAGATGAGCGTGTGTGTGTGTGTGTGTGTGTGT 2386
 QY 2230 GCTTGTCTATGATGAGGAGATGTTGGTTCAAGTATGATCAGAGCCAAATGCTGTG 2289
 Db 2387 TCGTGCAGAAATGATGCGGTTAAAGTGAATTTCTTTATG---AGGACAAATGCGCAGA 2443
 QY 2290 GTTCAAGTCCATGGCCAAATGATCTTCATCG 2222
 Db 2444 GTTCAAAATCTGTGGCTAGTTCTTCACATCG 2476
 RESULT 10
 AA21001
 ID AA21001 standard; cDNA to mRNA; 2746 BP.
 XX
 AC AA21001;
 XX
 AC
 XX
 DT 26-OCT-1999 (first entry)
 XX
 DE Nucleic acid encoding a raffinose synthase protein.
 XX
 KM Raffinose synthase; plant; broad bean; sucrose; raffinose; ss.
 XX
 OS Vicia faba.
 XX
 Key Location/Qualifiers
 FH 101..2500
 FT /*tag= a
 FT /product= "raffinose synthase"
 XX
 PN JF11215984-A.
 XX
 PD 10-AUG-1999.
 XX
 PF 12-DEC-1997; 97JP-0342899.
 XX
 PR 28-NOV-1997; 97JP-0329006.
 XX
 PR 18-DEC-1996; 96JP-0338673.
 XX
 PA (SUMO) SUMITOMO CHEM CO LTD.
 DR WPI: 1999-511112/43.
 DR P-PSDB: AA130142.
 PS New raffinose synthase gene - is prepared from a plant material
 Claim 6: Page 21-24; 40pp; Japanese.
 CC The present sequence encodes a raffinose synthase protein. The
 CC sequence is isolated from plant material of broad beans. The
 CC protein forms raffinose by complexing alpha(1 to 6)- D-galactosyl
 CC hydroxyl group of the 6c of D-glucose residue in sucrose molecules.
 SQ Sequence 2746 BP; 780 A; 552 C; 621 G; 793 T; 0 other:

Query Match	40.7%	Score 956.6;	DB 20;	Length 2746;
Best Local Similarity	66.6%;	Pred. No. 3.5e-297;		
Matches 1421; Conservative	0;	Mismatches 694;	Indels 18;	Gaps 3

Oy	193	GGTTGCTTTGTGGATTTGCGATCGCGGTGGAAACCGMTATAGCCGACATGTTTTCGATTGGG	252
Db	359	GGTTGTTTCTGTTGGTTTCAACTCACCAGAACCCAAAGCCACACAGTATGTTTCCACTGGCC	418
Oy	253	AAGCTGAGGATATTTGCGTTTATGACTATTTTCAGGTTTAAAGTTTGGTGACTACACAC	312
Db	419	AAATCTAAAAGGAATCAAAATTCATGACACATATTCGGTTTCAAAAGTTTGGTGACAACATCAC	478
Oy	313	TGGGTTGGTGGAANAGTGGGGAATCTTGAAATGGAGACTCAGATTGTATCTCTTGACAG	372
Db	479	TGGGTCGGAACAAATGACAGCAGATACACACGAAACCAAAATGTTATATCTTGACACAA	538
Oy	373	TCGATATTC---TGGTGACCGCTATGTTTTCCTTCTTCGATCGTTGAGAGGACCGTTCCGA	429
Db	539	AAGCAGTCCGCGAGACGACCTTATGTCTTACGTCTCCCATCTCTGAAAACACCTTCGGA	598
Oy	430	ACCTTCGATTCAAGCCTGGGGAATGATACATTTTTCGATGTTTGTCTGACAGTGTTTCGTG	489
Db	599	ACCTCAGCTCCAAACCCGCTCCAAACGATTCACATAGCCATGCTCCGTGGAAGGCGTTTCAACA	658
Oy	490	AAATCTTTGATGCATCGTTCCGAAATATGTTATGTTATCTTACATCGTGAGATACCGGTTT	549
Db	659	CATGTCCACCGGCTCAAGCTTCAAAAGCATGTCTTTTACATCCATCTCAGTAAACAGCCCATAC	718
Oy	550	GCACATTGTTAAAGAGCGGATGAAGATCGTAGAGACCCATCTTGAACATTTTCGCTTGTG	609
Db	719	AGTATCTTAAAGAAAGACAGATTAAAGTAATCCAAACTCAGATTAGSAAATTCACAACTCTT	778
Oy	610	GAGGAGAAAGCTCCACAGGTATCGTGACAAATTCGGTTGGTGACGTGGAGCGGCTT	669
Db	779	GAAACAAAACAGACCTATATATTATAGCAAAATTCGGTTGGTGACGTGGAGATGCTTTT	838
Oy	670	TACCTAAGGTTTCATCCACAGGGCGTAAATAGAAGCGTGAGGCATCTGTGACAGGCGGT	729
Db	839	TACTTGAAGGTTTCATCCAAAGGTGATGAAGAGGTAAAGTCTCTCACAGATGTGTGT	898
Oy	730	TGTGCTCCCGGTTTATGTCTTAATCAGCAGATGTTGGCAATTCATCGGACACAGATTGGAT	789
Db	899	TGTCTCTCCCGTTTCGTCATTAATCGACAGCGGTGGCAATTCATTGTTCATAGCAGATAC	958
Oy	790	CCCATCTACCAAGAGGAATGAACCAACCGTCGTGCGCGGAGCAAAATCCCTGCGCTT	849
Db	959	GATGAAATGATTTCAGSAAATGAACCGCAACCTCAGCGGGGAAACAAATGCATGCAAGCTT	1018
Oy	850	TTGAAATTCACAGAGATTACAAATTCGCTGACTACGTCAATCCCAAGGCCACCGGCCCC	909
Db	1019	GTTAAATATCAGCAAGCAATTTCTAAGTTTAAAGATATGAAATCCGTA-----A	1066
Oy	910	CGACCCGCGCAGAAAGGGGATGAAGGCGTTTATAGATGAACCTCAAGAGAGATTAAAGCT	969
Db	1067	AATGAGAGGAAGAAAGSTTTGGGTGTTTGTGTAGAGGATTTTAAGAAAGATTTGGGAGT	1126
Oy	970	GTCGAGCAGTGTATGTTTGGCATGCTTTGTGTGATATTGGGATGGCCTTTCGCCCGAG	1029
Db	1127	GTCGAGATGTATTATGTTTGGCATGCTGCTTTGTGGGATTTGGGCGGGGATTAGCGTCGA	1186
Oy	1030	GTCGCTGCGTTCCTAGAGCACTGTGATTCAGCCAGTCCTTTCACAGGCGTCAAGATG	1089
Db	1187	GTCGATGGAGTCCGAAAGCTATAGGTTGTGTTGTTCCGAAGTGTCTCAAGGGTTTGAAGATG	1246
Oy	1090	ACGATGAGAGATTTGGCGGTGATGAATGTTCTTCAATAAAGTGGGCGCTGTCTCCGCG	1149
Db	1247	ACGATGAGAGATTTGGCGGTGATGAATGTTTGAAGAACGGTGTGGGCTATGTGCGCGCA	1306
Oy	1150	GAGAAAGCTGAGAGATGTACGAAGACATTCACTCATTTTGGAAAAAGTTGGGATGCA	1209
Db	1307	GATTTTCCACATGAGATGTTGATGGGCTTCACTCTCATTTTGGGATTCGCGCGGAATTCAC	1366

QY	1210	GGTGTAAAGATTGACGCTTATTCACACCTATTGGAGATGTTGTGAAAGACTATTGAGGGAGA	1209
Db	1367	GGGTGTAAAGTGTGATTGTATTCACATCTGGCTTGGATTACTATCAGAGAAATTGTTGAGCA	1426
QY	1270	GTTGATTTGGCAAAAGCATTTATCAAAAGCAATGACCAAATCAATTAATTAACATTTTAA	1329
Db	1427	GTTAGAGCTAGCAGAGACTTATTTCAAAGCACTTACCTCATCTAGTAAAGAAATTTTCAA	1486
QY	1330	GGAATGGAGTCAATTGGCAAGTATGAAACAATTGTAAACGACTTCATGTTCCTTGGCAGGAA	1389
Db	1487	GGCAATGGTGAATTTGCTAGCATGGAGCAATGGAACGACTTCTTCTCTCGCCAGCAGAA	1546
QY	1390	GCTATCTCTTGGTGGTGTGGTGATGACTTTGGTGGCAGGAGCCCTCTGGTGATCCA	1449
Db	1547	GGCATATCCCTCGGCCCGCCGTCGGAGATGATTTTGGTGGCTCTGATCACTATCTGTGATCCA	1606
QY	1450	AAGCGTACGTTTGGGCTCCAAAGATGTCACATGGTTCATTTGGCCACAGCAGCTGTGG	1509
Db	1607	AATGGTACATATTGGCTTCCAAGTTGTTCACATGGTACATTTTGGCTTACACACGTTTATGG	1666
QY	1510	ATGGGGAACCTTCATCCACCCCTGACGGGATATGTTCCAAATCCACCCACCCTTGGCCGCC	1589
Db	1667	ATGGGAATTTTCATTCAGCCACAGATTTGGAGCATGTTTCACTCCACATCATCTTGTGTGAA	1726
QY	1570	TTTCATGCTGCTCTCGAGCCATCTCTGGTGGCCGAGTCATGTTAGTATTTGTGTGGGA	1629
Db	1727	TTTTCATCCCGCTTCACAGAGCCATATCCGGCGGACCAATTTATGTTAGTATTTGTGTGGT	1786
QY	1630	AAGCATAACTTTGATCTCTTGAAAAAACTAATGTCCTCTGATGATTCGATCTCTTCGAAGT	1689
Db	1787	AATACAAATTTCAAGTTGCTCAAAATCTCTTGGTTTGGCCCGAATGTTCTATCTTGGGTTGT	1846
QY	1690	GAGTACTATGCACTCCGCACTCGCATGTTGTTTGGTGAAGACCTTTTGCATTAATGGAGAA	1749
Db	1847	CAACATTACGCACTCCCTACAAAGATTTGCTTGTGAAAGACCTTTGTCATTAATGGCAA	1906
QY	1750	ACTATGCTTAAGATTGGAAATCTCAACAAGTTCACGTGAGTATTTGTGATCTCAATGTC	1809
Db	1907	ACAATGCTGAAAAATTTTGGATCTCAACAAATATACAGGTGTTTGGGCTTTTCAACTGC	1966
QY	1810	CAAGAGAGAGATGCTGCTGTGAGACACCCCGCAACCAATGCTTTTTCACAAATATCAACAA	1889
Db	1967	CAAGTGTGTGGTGTGTCTGTGAGCAGCGGCAACAAAGTATCTGCAATTTTCAACGC	2026
QY	1870	CGAGTACATCCAAAACTCAACCCCAAAAGCAATTAAGATGGCAGTGGAGGAAAAACCTATC	1929
Db	2027	GGGTGACATGTTATTCGAACGTCCCGAAGACATTTGAATGTTGCAATTTGGCAAACTCCAAATG	2086
QY	1930	TCATATGAAGCGTTAAACCTTTGGGCTTTTACCTCTATCAAGCCAAAAACCTTATCTC	1989
Db	2087	AGCACCAAAAGGTGTGATTTTGTGCTGTGTATTTTTCAAAGAAAGAAATTTGAGCTC	2146
QY	1990	TCCAAGCCCTCTCAAGATCTTTGACATTAAGCTCTTGACCCATTTGCAATTTGGAGTCTATCACT	2049
Db	2147	ATGAAGTGTCTGTGATTAATTTGAAGAAATTTTGGCTTGACCACTTTTATGTTTGAAGCTAATGCA	2206
QY	2050	GTTTTCACCAAGTGAACAACTATCCAAAGCTCTCTACATCTTTGGCCCAATGGGCTGTG	2109
Db	2207	GTTGCTCCACGTGAAGATGTTTTCGAAAAAGTTTATACATTTTTCGACCCGATTTGGGTTA	2266
QY	2110	AACATGCTTAACACTACTGTGAGGCATTCACATCTGTGACTATGACGATGACCTTAAGTCA	2189
Db	2267	AACATGCTGAACCTGTGTGTGGGATTCAGATCTCTGTGACTTTCATGATTAATGCAAACTTG	2386
QY	2170	GTCGAGATTGCTGTCAAAAGGCTGTGTGAGATGCGAGTATTTTGCAATCGAAAAACCAAG	2229
Db	2327	GTCAGATTTGGGAGTGAAGGCTTCCGGGAGATGAGCGTGTTCGCTGTAGCAAAACCGGTT	2386
QY	2230	GCTTGTGATTAATGATGGGAGGAGATGTTGGGTTCAAGTATGATCAGAGCAACAAATGTTGTG	2289
Db	2387	TGCTGCAAAATTTGATGGGGTTTAAGGTAAATTTCTTTATG---AGCACAAAATGGCAAGA	2443
QY	2290	GTTCAAGTGGCAATGGCAATTTGATTTCTTCATCG	2322

```

Db      2444 GTTCAAAATTCGTGGCGCTAGTCTTCACACATTTG 2476
||||| | | | | | | | | | | | | | | | | | | | | |
RESULT 11
ID      AAD00335
        AAD00335 standard; cDNA: 2524 BP.
XX      AAD00335;
XX      AAD00335;
XX      09-AUG-2000 (first entry)
XX      Soybean raffinose synthase cDNA from clone sfl1.pk125.d4.
XX      Soybean; raffinose synthase; raffinose saccharide;
XX      clone sfl1.pk125.d4; nutritional; soy protein; ss.
XX      Glycine max.
XX      Key Location/Qualifiers
XX      CDS 25..2301
XX      FT /*tag=a
XX      FT /product="Raffinose synthase"
XX      PM WO200024915-A2.
XX      PD 04-MAY-2000.
XX      PF 22-OCT-1999; 99WO-US24923.
XX      PR 23-OCT-1998; 98US-0105451.
XX      PA (DUP0 ) DU PONT DE NEMOURS & CO E. I.
XX      PI Allen SM, Hiltz WD;
XX      DR MPI: 2000-350754/30.
XX      DR P-PSDB; AAY70978.
XX      PT Nucleic acids and encoded proteins involved in the biosynthesis of
XX      raffinose, useful for producing soybean seeds with a reduced raffinose
XX      content and therefore improved nutritional quality -
XX      PS Claim 2; Page 46; 58pp; English.
XX      CC The present sequence is a cDNA encoding raffinose synthase from
XX      clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
XX      library sfl1. Raffinose synthase is involved in the biosynthesis
XX      of raffinose and higher homologues in the raffinose saccharide family
XX      from sucrose. The present sequence is useful for reducing the raffinose
XX      saccharide content of soybean seeds which improves the nutritional
XX      quality of the soy protein products derived from them.
XX      CC
XX      SQ Sequence 2524 BP; 671 A; 556 C; 679 G; 618 T; 0 other;

Query Match 40.3%; Score 948.6; DB 21; Length 2524;
Best Local Similarity 67.3%; Pred. No. 1.2e-294;
Matches 1436; Conservative 0; Mismatches 649; Indels 48; Gaps 5;

OY      193 GGTTCCTTTGGTGAATTCAGCGCTGGAACCTGATAGCCGACATGTTTGCATTTGGG 252
        || ||||| | | | | | | | | | | | | | | | | | | | | | |
DB      190 GATGCTCTTCCTGGTTTCACGCCACCTCCCAAAATCCCGCAGCTGGCACCCTTAGGA 249
OY      253 AAGCTGAAGATATTCGTTTATGAGTATTTTCAGGTTTAAAGTTTGGTGAATACAC 312
        ||||| | | | | | | | | | | | | | | | | | | | | | |
DB      250 CAGCTTAAAAACATTAAGCTTCACTTCATCTCCGCTCAAGGTTTGGTGACCACTTC 309
OY      313 TGGGTGGTGCAGAAATGCTGGGATCTTGAATGAGACATCAATTTGATCTTGAAGA 372
        || | | | | | | | | | | | | | | | | | | | | | | |
DB      310 TGGAGTGGCTCCACAGCGCGCGACCTGGAACCGAAACCAATTCCTCATGTGCCA---- 365
OY      373 TCAGATTCGTGCGAACCGGATGTTTCTTTCGATGCTTGAGGAGCGTTCCGAAAC 432
        | | | | | | | | | | | | | | | | | | | | | | |

```

```

DB      366 -----ATCCACCCTTATGTTCTTCTTCCTAACCATCTCCAAACCCCATTTGCGGC 417
OY      433 TCGATTCAGCCTGGGGATGATGATCTTGTGATGTTTGTGTGAGAGTGTTCGTGAA 492
        || | | | | | | | | | | | | | | | | | | | | | |
DB      418 TCGCTCAGCCTCATTACAGACGACCAACGTTGGCGTGTGTGTGAGAGCGGCTCAGCCAC 477
OY      493 GTTGTGATGATCGTTCCGAACTATGTTGTATCTTCATGCTGTGTATGATCATCGTTTGA 552
        || | | | | | | | | | | | | | | | | | | | | | |
DB      478 GTACAGCCCTCATCATCTGACACTGTCTACTTCCAGCAGGAGCAACCCCTTCAG 537
OY      553 CTGTGTTAAAGAGGCGATGAAGATCGTAGAGACCATCTTGGAACTTTTGGCTTTGGAG 612
        || | | | | | | | | | | | | | | | | | | | | | |
DB      538 CTGCTCAAGAGAACCATTCGCGCTGTCGCGGCGCTTGGGAGCTTCAAGCTTCTGGAA 597
OY      613 GAGAAAGCTCCACCAAGTATCGTAGCAAAATTCGTTGTGTGACGCTGGAGAGCGGTTTAC 672
        || | | | | | | | | | | | | | | | | | | | | | |
DB      598 GAGAAACAGTTCCGGGAATGTGTGACAAAGTTGGTGTGTGACGTGGAGCGCTTTTAC 657
OY      673 CTACCGTTTCATCCACAGGCGCTAATAGAAAGCGTGAAGCATCTGCTGCAGCGCGTTGT 732
        || | | | | | | | | | | | | | | | | | | | | | |
DB      658 CTGACGCTCAGCCCTGAGGCGCTCAGAGAGGCGTGAAGGCGCTGTTGACGCGGTTGT 717
OY      733 CCTCCCGGTTTAACTCTTAATCGACGATGTGTGCAATTCATCGACAGCATTTGCAATCCC 792
        || | | | | | | | | | | | | | | | | | | | | | |
DB      718 CTTCCGCGATTCGTTCTGATCGACGACGCGCTGCAATGCAATCCAGATTCGCAATCC 776
OY      793 ATCCCAAGAAAGAAATGAACCAACCGTCCGCGGAGCAAAATGCCCTGCTTTTGG 852
        || | | | | | | | | | | | | | | | | | | | | | |
DB      777 --GAGAAAGAGGAGATGATATGACGCTGTGCGGAGCAAAATGCCCTTGTGAT 834
OY      853 AATTTCCAGAGATTTACAAATTCCTGTGACTACGTCAATCCCAAGGCCACGCGCCCGCA 912
        | | | | | | | | | | | | | | | | | | | | | |
DB      835 AGTTACGAGAGAAATTAACAAGTTTAGAGCTA-----T 867
OY      913 GCGGCGCAAGAGGAGATGAAGGCGTTTATAGATCACTCAAGAGAGTTTAAGACTGTG 972
        || | | | | | | | | | | | | | | | | | | | | | |
DB      868 AAGAGAGGAGAGGCGTTGAAGGGGTTTGTGAGGAATTAAGAGAGAGATTTGGGTGGTG 927
OY      973 GAGCATGTTTATGTTTGGCATGCTTGTGTGATTAATGGGTGGCTTGGCCCGCAGGTG 1032
        || | | | | | | | | | | | | | | | | | | | | | |
DB      928 GAGTACGTTACCTGTCGACGCGCTGTGCGGATTAATGGGAGGCGTGGAGCGGTG 987
OY      1033 CCTGCTTGCCTGAGGACGCTGTGATTCAGCCGCTCTTCAACGAGGCGTGCAGATGAG 1092
        || | | | | | | | | | | | | | | | | | | | | | |
DB      988 GCGGGATGGCGGAGGCGCGCGGTGAGAAAGCCAAACCTGACGAGGCGTTGAAGGGAGAG 1047
OY      1093 ATGAGAGATTTGGCGGTGATTAAGTGTCTTCAATAGCTCGGCGTGTCCCGCGAG 1152
        || | | | | | | | | | | | | | | | | | | | | | |
DB      1048 ATGAGAGATCTGCGGTGAGCAAGATTTGAAATATGGGGTGGGTGGCGCGAG 1107
OY      1153 AAGCTGAGAGATGTACGAGAGACTTCATGCTCATTTGGAAAAAGTTGGATCGACG 1212
        || | | | | | | | | | | | | | | | | | | | | | |
DB      1108 CTGGTGGGGAAATGTATAGAGGCTTCAACGCGCTTGGAGTGTGGATTTGGG 1167
OY      1213 GTTAAGATGAGCTTATCCACCTATTGAGAGATGTGTGGAAGACTATGAGAGAGAGTG 1272
        || | | | | | | | | | | | | | | | | | | | | | |
DB      1168 GTCAAGTTGATGTATCTCACTTGTAGAAATGTGTGTGAGAAATATGAGAGCGAGTG 1227
OY      1273 GATTTCGCAAGGCAATATTACAAAGCAATGACCAATGCAATAAATTAAGATTTAAAGA 1332
        || | | | | | | | | | | | | | | | | | | | | | |
DB      1228 GATATGCGCAAGACATATTACAAAGCTTCACCTGCTCCGTGAGAAACATTTTAAGGCG 1287
OY      1333 AATGAGATATTGCAAGTATGCAACATTTGAACGACTTATGTTCTTGGACGAGAGCT 1392
        || | | | | | | | | | | | | | | | | | | | | | |
DB      1288 AACGGCGTATTGCCAGCATGTGAGAGCAATTCAGATTTGCTGTGGGAATCAAGA 1347
OY      1393 ATCTCTCTGTGCGTGTGTGTGATGACTTTTGTGTGACGAGACCCCTGTGGTATCCAAAC 1452
        || | | | | | | | | | | | | | | | | | | | | | |
DB      1348 ATATCCCTTGTGCTGTGTGTGGGATGATTTCTGTGTGACATGACCTTATATGTATCCAAAT 1407
OY      1453 GGTACGTTTGGCTCCAGAGATGTCAACATGTTTGTGCAACAGCAAGCTTGTGATG 1512
        || | | | | | | | | | | | | | | | | | | | | | |
DB      1408 GTTACATTTTGGTACAAAGGTGTCAATGTGTGATTTGTGCAATCAACAGCAAGCTTGTGATG 1467

```


Db 421 GGTGTTGATCGACGAGGGTGGCATTCATTGTCACGACACGAGCGCTCACCCGA 1480
 QY 804 AGGAATGAACCAACCCGTCGCGGAGCAAAATGCCCCGCTTTTGAATTCACAGA 863
 Db 481 GGGATGGGAGAACCTCCGCCGAGAGCAAAATGCCCTGAGGTGATCAAGTTTGAGGA 540
 QY 864 GAATTCACAAATTCGTCGACGTCAATCCCAAGGCCACCGCCCCGAGCGGCCAGAA 923
 Db 541 GAATTCACAAATTCGTCGACGTCAATCCCAAGGCCACCGCCCCGAGCGGCCAGAA 923
 QY 924 GGGAGTGAAGCGCTTTATAGATGACATCAAGAGAGTTTAAGACGTGGAGCATGTTA 983
 Db 598 GGGAGTGGGGGCTTTATTCGTGACATGAAGCAATTTTAAAGAGTGGACATCGTGA 657
 QY 984 TGTGTCGATGCTTTGTCGATATTTGGGGTGGCTTCGCCCGCAGGTGCTTGGCC 1043
 Db 658 CGTGTGTCATGCTGTTGTGTTATTTGGGGCGGGCTCAGGCCCAATGTTCCGGGCGTCC 717
 QY 1044 TGAAGCAGCTGTGATTCAGCAGTGCCTTTCACAGGCGCTGACATGACATGAGGATTT 1103
 Db 718 CGAGGCTAAGCTCATTGAGCCCAAACTGACTCCTGGGCTTAAACACACCATGAGAGATTT 777
 QY 1104 GGGCGTGAATGAATGTTCTTCTTAAGTCCGGGCTGGTCCCGCGAGAGGCTGAGGA 1163
 Db 778 GCGTGTGATGAATGATGTCAACAAATGGCTGGCTGTGTCACCGAGTGTGTTGACA 837
 QY 1164 GATGTACGAGAGCTTCATGCTCATTTTGGAAAAAGTTGGAGTGCAGCGTGAAGATTGA 1223
 Db 838 AATGTATGAAGATTCATTACATCTCGAATCTGTGGGATTTGATGAGATCAAGATTGA 897
 QY 1224 CGTTATCCACATTTGGACATGTTGTGTGAAGCTATGAGAGGAGTGGATTTGGCAA 1283
 Db 898 CGTATTCACATTTGTTGAAATGTTGTGTAAGACTATGAGGAGAGTGGACTTAAACCA 957
 QY 1284 GGCATATTCACAAAGCAATGACCAATCAATTAATAAATTTTAAAGAAATGAGTCAAT 1343
 Db 958 GGGTATTTCAAGGCTTTATCAAGCTCAGTATTAACAACCACTTAACGGCAAGCGCTCAT 1017
 QY 1344 TGCAGTATGGAACATTTGACAGCTTCATGTTCCCTGGCAGGAAGCTATCTCTTGG 1403
 Db 1018 CCGTGGCCCTGGAGCACATGATGATGTTTCTCGGAACCGAGCCATTAACCTTGGC 1077
 QY 1404 TCGTGTGGGATGATCTTTGGTGGTGGACGGCCCTGGATGATCCAAAGGTTGCTTTG 1463
 Db 1078 TCGTGTGGGATGATCTTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1137
 QY 1464 GCTCCAGAGATGTCACATGTTTCATTTGTCACACGACAGCTTGTGATGGAGTAATTCAT 1523
 Db 1138 GTTGCAAGGTTGACATGTTGTCATGTCGCTTACACAGACATGATGATGATTTTCAT 1197
 QY 1524 CCACCTGACTGGGATATGTTCCATCAACCAACCCCTTGTCCGCTTCATGCTCCCTC 1583
 Db 1198 CCACCTGATGAGCATGTTCAATCGACTCACCTTGGCGTGAATTCACACCTCCCTC 1257
 QY 1584 TGAAGCATCTGCTGGTGGCCGATGATGTTAGTGTGTTGTTGGGAAACATTAACCTTGA 1643
 Db 1258 ACAGGCATCTCCGGGGGCCCATTTACATGCTGACTGCTGGTGGGAAACACACACTTCA 1317
 QY 1644 TCTTTCGAAAAAATAGTCTTCTCTGATGATGATGATGATGATGATGATGATGATGATGAT 1703
 Db 1318 GCTCCTTAGAGCTGCTCTTCCGATGCTCCATGCTCCCTGTTGATTTACTACGCGCT 1377
 QY 1704 CCCGACTCGCATGTTGTTTGAAGACCCCTTGTGATTAATGAGAACTATGCTTAAGAT 1763
 Db 1378 TCCGACTCCGATGCTCTTGAAGATGACACTTCAATGGCAAGACATATGCTCAAAAT 1437
 QY 1764 TTGGAATTCACAAAGTTCACGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1823
 Db 1438 TTGGAATTCACAAAGTTCACGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1497
 QY 1824 GTGTGTCGATGACGCGGCAACCAATGCTTTTACAAATACTCAAAACGATGATCAACCA 1883
 Db 1498 GAGCGCGGAGAGTGGCTGCGAACCAATGCGCTGCGGAGTATTCGCCAGCGCTCTCTCTAG 1557

QY 1884 AACTAACCAAAAGACATAGATGACACAGTGGAGAAAACCTTATCTTATGAGGCGT 1943
 Db 1558 CGGTGGTCGAGTACATGATGATGAGAAAGCAAGACAGATCCGATGACGTCGAGCGCT 1617
 QY 1944 TAAACCTTTGGCGCTTTACCTTATATCAAGCCAAAAAATTTATCTCTCAAGCCCTCTCA 2003
 Db 1618 CAAAACATTTGCGCTTGTACCTATTCACGAGAAAGAACTGCTCTTCTTAAGCATCAGA 1677
 QY 2004 AGATCTTGACATGCTCTTACCATTCGAAATTCGACCTCATCAGTGTTCACACAGTAC 2063
 Db 1678 CAAAATGACATGACGCTTGAAGCCCTTTCGATTTTGAGCTGATTAACGTTTCTCAGTCAA 1737
 QY 2064 CAAACT 2069
 Db 1738 AACTCT 1743

RESULT 13

AA210003
ID AA210003 standard; cDNA to mRNA: 1762 BP.

AC AA210003;
 DT 26-OCT-1999 (first entry)
 DE Nucleic acid encoding a raffinose synthase protein.
 XX
 KW Raffinose synthase; plant; sucrose; raffinose; ss.
 XX
 OS Stachys sieboldii.

FH Key Location/Qualifiers
 FT CDS 2..1762
 FT /tag= a
 FT /product= "raffinose synthase"
 FT /note= "no termination codon given"

PN JPI1215984-A.
 PD 10-AUG-1999.
 XX
 PF 12-DEC-1997; 97JP-0342899.
 XX
 PR 28-NOV-1997; 97JP-0329006.
 PR 18-DEC-1996; 96JP-0338673.
 XX
 PA (SDMO) SUMITOMO CHEM CO LTD.
 XX
 DR WPI: 1999-511112/43.
 DR P-PSDB: AAY30144.
 XX

PT New raffinose synthase gene - is prepared from a plant material
 XX
 PS Claim 13: Page 31-34: 40pp: Japanese.
 XX

CC The present sequence encodes a raffinose synthase protein. The
 CC sequence is isolated from plant material. The protein forms raffinose
 CC by complexing alpha(1 to 6)-D-galactosyl hydroxyl group of the 6C
 CC of D-glucose residue in sucrose molecules.
 XX

SQ Sequence 1762 BP; 395 A; 451 C; 499 G; 417 T; 0 other;

Query Match 37.7%; Score 886.8; DB 20; Length 1762;

Best Local Similarity 70.0%; Pred. No. 8.7e-275;

Matches 1223; Conservative 0; Mismatches 517; Indels 6; Gaps 2;

QY 324 AATGTTGGGATCTTGAATCGAGACTCAGATTGTGATCTTGAAGACTCAGATTCTGG 383
 Db 4 AAACGGGTGCGATCTGACCGGGAACCTCAATAGTGTGTCGCAAGTCCGA--CGA 60
 QY 384 TGAACGCTATGTTTCTCTCTTCCGATCGTTGAGGAGCCGTTCCGAACCTCGATTGAGCC 443

Db 61 CAGGCCCATACATGCTGCTTCCGCTCATCGAGGGGCAATTTCGGGCTTCCCTTCAGCC 120
Oy 444 TGGGATGATGATCTTGTGATGTTGTGTCGAGAGTGGTGTGTCGAAGTTGTTGATGC 503
Db 121 CGGTGGATGATGATTTTATGATATTTGTCTCGAAGCGGGGTCAACCAAGTCAACAGATC 180
Oy 504 ATGCTTCCGAAGTATGTTTATCTTTCATCTGATGATCCGTTTTCATCTGTTTAAAGA 563
Db 181 CTGCTTCCGCTGCTTCTACATGACGCGGTGATGACCTTTTACCCTGGTGAAGGA 240
Oy 564 GGGGATGAAGATGCTGAGACCCATCTTGGAACTTTCCGTTGTGGAGAGAAGACTGC 623
Db 241 CCGCGGAGAGGGGCGCGCCACCACTTCGGGAGCTTCAGGCTGCTGAGAGAGAACTCC 300
Oy 624 ACCAGATCTGAGCAAAATTCGGTTGTGTCAGCTGGGAGCGGCTTTTACCTTAAGGTTCA 683
Db 301 GCGGGGATCTGTCGACAAATTCGGGTGTGTCACGTGGATGCTTCTTACCTCACTTCCA 360
Oy 684 TCCACAGGGCGTAATGAGAGGCGTGAAGCATCTCTGACAGCGGCGTTTCTCCGCTTT 743
Db 361 GCGCCAGCGGCTTATGAGAGGCGGTGCAAGGGCGTGTGAAGGGGATGTCGCGCGGGGCT 420
Oy 744 AGTCTTAATCGAGATGTTGGCAATCCATCGGACAGATTCGATCCATCAACCAAGA 803
Db 421 GGTGTGATCGACGAGGGGTGCGAGTCCATTTGTTCAGACAAACGACGCGCTCACCCGA 480
Oy 804 AGGAATGAACCAACCGGTGCGCGGAGCAAAATGCCCTGCCCTTTTGAATTTCCAAAG 863
Db 481 GGGGATGGGAGAACCTCCGCGGAGAGCAAAATGCCCTGAGTTTATCAAGTTTGAAGA 540
Oy 864 GAATTAACAATTCCTGACTACGTCATCCAGAGGCCACCGGCGCCGAGAGCGCGGAGAA 923
Db 541 GAATTACAAATTTCAGGAGAGACAGAGAGCCGAATTAACCTGGGCGC---GGGCGCGAATAC 597
Oy 924 GGGGATGAAGCGCTTATAGATGAACTCAAGAGAGATTGAAGCTGTGAGAGATTTTA 963
Db 598 GGGGATGGGCGCTTATTCGTGATCAATGAAGCAATTTCAAAAGTGTGACATCACTTA 657
Oy 984 TGTGGCATGCTTGTGTGATATTTGGGTGCGCTTCGCGCGCAGGTGCTGCTTGGCC 1043
Db 658 CGTGTGCAATGCTTGTGTGATATTTGGGTGCGCGCTCAGGCGCAATGTTCCGCGCTGCC 717
Oy 1044 TGAAGCAGCTGTGATTCACCCAGTCTTTCACAGGGGCTGACAGATGAGTGAAGATTT 1103
Db 718 CGAGGCTAAGCTATTCAGCCCAACTGACTCTCGGGCTTAAAGACCACTATGAAATTT 777
Oy 1104 GGGGATGAATGATGTTCTTCTCATAAAGTCCGGCTGTGCTCCCGCGAGAGCTGAGGA 1163
Db 778 GCGTGTGATTAAGATTTGCAACAATGCGGTGTGCTGCTCCACCGAGTTGTTGAACA 837
Oy 1164 GATGTGAAGAGACTTCATGCTCATTTTGAAGAAAGTTGGATGAGCGTGTGAAGTTGA 1223
Db 838 AATGTATGAAGATTTACATTCATCTCGAATCTGTGGGATTTGATGATCAAAATTTGA 897
Oy 1224 CGTTATCAACCTTATGAGATGTTGTGTAAGACTATGAAGAGAGATGATTTGGCAAA 1283
Db 898 CGTCAATCAATTTGTGGAATGTTGTGAAGACTATGGGAGAGATGACTTATGCCAA 957
Oy 1284 GGCATTTTCAAAAGCAATGCAACCAATCAATTAATTAACATTTTAAAGGAATGAGTCAAT 1343
Db 958 GCGTTATTAACAAGGCTTATCAAGCTCAGTTAAACAACCACTTCAACGCGCAT 1017
Oy 1344 TGCAGATATGAAGCATTTGACGACTTATGTTCTTGGACGAGAGATATCTCTTGG 1403
Db 1018 CCGTGGCTGTGAGACATTCGATCTTCAATGTTTCTCGAAGCGAGGCCATTCCTTGG 1077
Oy 1404 TCGTGTGATGATGATCTTGTGTCAGGAGCCCTCTGTGATCAACAGGTAAGTTTG 1463
Db 1078 TCGTGTGCGGAGATGATTTTGTGCTACTATCATCTGGAAGATCCCAATGGCAGCTTCTG 1137
Oy 1464 GCTCCAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1523
Db 1138 GTTCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1197

Oy 1524 CCACCTGACTGGGATATGTTCCATCCACACCCCTTGTGCGGCTTCCATGCTGCTC 1583
Db 1198 CCACCTGATTTGGAGATATGTTTCAATTCAGCTACCTTGGCTGATTTCCAGCTGCTC 1257
Oy 1584 TCGAGCATCTCTGTGAGCGCCGATCTATGTTAGTGAATCTGTGGAAAGCATATTTGA 1643
Db 1258 ACAGCATCTCTCGGCGGCGCCATTTAGTCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1317
Oy 1644 TCTTGTGAAGAACTAGTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1703
Db 1318 GCTCTTGAAGAGCTGCTGCTTCTGATGATGATGATGATGATGATGATGATGATGAT 1377
Oy 1704 CCGCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1763
Db 1378 TCGCATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1437
Oy 1764 TTTGAATCTCAACAGTTCATGAGTGAATGTTGTCATTCATTCATTCATTCATTCAT 1823
Db 1438 TTTGAATTTATCAAGATTCACCGAGTTGTGGAATTTCAACTGCGCAAGTGGCGGTG 1497
Oy 1824 GTTCTGTGAGACAGCGCGCAACCATGCTTTTACATCTCAATCTCAATCTCAATCTCA 1883
Db 1498 GAGCCCGGAGAGTCCGTGCAACCAATGCGTCCGATATTTCCAGCGGCTCTCTAG 1557
Oy 1884 AACTAACCCAAAGACATGAAATGAGCAGAGTGAAGAAACCCATCTTATTAAGGCGT 1943
Db 1558 CGTGTGCTGAGTACATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1617
Oy 1944 TAAACCTTTGCGCTTACCTCTATCAAGCCAAAGAACTTATCTCTCAAGCGCTCTCA 2003
Db 1618 CAAACATTCGCTGTTTACCTATTCACGAGAGAAAGTCTCTTCTTCTTCTTCTTCT 1677
Oy 2004 AGATCTTGAATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2063
Db 1678 CAAATTCGATCAGCTTGTGAGCTTGTGATGATGATGATGATGATGATGATGATGAT 1737
Oy 2064 CAAACT 2069
Db 1738 AACTCT 1743

RESULT 14
AA220210
ID AA220210 standard: cdna: 1762 bp.
XX
AC AA220210;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rapeseed raffinose synthase cdna.
XX
KW Raffinose synthase: rapeseed; transgenic plant; ss.
XX
OS Brassica napus.
XX
FH Key 1.1719 Location/Qualifiers
FT CDS /tag= a
FT FT /note= "this region is specifically claimed in
Claim 9"
PD EP953643-A2.
XX
XX 03-NOV-1999.
XX
XX 27-APR-1999; 99BP-0107430.
XX
XX 30-APR-1998; 98BP-0120550.
XX 30-APR-1998; 98BP-0120551.
XX 04-DEC-1998; 98BP-0345590.
XX 10-DEC-1998; 98BP-0351246.
XX

PA (SUNO) SUMITOMO CHEM CO LTD.
XX Watanabe E, Oeda K;
PI
XX MPI: 1999-593144/51.
DR P-PSDB; AAY32075.
XX
PT New sense and antisense genes, useful for altering the level of
PS raffinose in food plants -
XX
PS Claim 9; Page 38-41; 55pp; English.

XX This is the nucleotide sequence of a rapeseed cDNA clone coding
CC for raffinose synthase (see AAY32075), a protein which can bind a
CC D-galactosyl group through an alpha(1-6) bond to the hydroxy group
CC attached to the carbon atom at the 6-position of the D-glucose
CC residue in a sucrose molecule to form raffinose. The cDNA was
CC isolated from rapeseed cv. Westar leaf cDNA by PCR. Probes or
CC PCR primers generated from plant raffinose synthase genes (see
CC AAY207-10) may be used to obtain other raffinose synthase genes
CC by labeled detection or amplification (claimed). These genes may
CC be used to control the levels of raffinose produced in plants.
CC Antisense genes can be used to knock out existing gene activity,
CC and sense genes to increase the level of gene activity. The
CC resulting transgenic plants may be used as a food source to alter
CC the growing conditions for gut enterobacteria, providing general
CC health advantages. The sense and antisense genes may also be used
CC in gene and phenotypic analysis of plants and for the selection of
CC plant clones with the desired characteristics with respect to
CC raffinose content.
XX

SQ Sequence 1762 BP; 435 A; 384 C; 459 G; 479 T; 5 other:

Query Match 36.8%; Score 866.4; DB 20; Length 1762;

Best Local Similarity 70.6%; Pred. No. 3.3e-268;

Matches 1200; Conservative 4; Mismatches 475; Indels 21; Gaps 3;

QY 607 TTGGAGGAGACACTCCACGATGCGGACAAATTGGTTGGTCACGTGGACGCG 666
DB 1 TTGGAGGAGGAAAAAGCGCGCGGAAATCGTGAATTCGGGTGGCGACATGGATGCG 60
QY 667 TTTTACCTAAGGTTTCATCCACGAGCGGTAATGGAAGGGGTGGGATCTCCGTCAGCGC 726
DB 61 TTTTATTATTCAGGTGACCTTGACGAGATTCATAGAGGTGTTAGTGTCTCGTCGACGG 120
QY 727 GGTGTCTCCCGGTTTAACTTCAATGACGATGTTGGCAATCCATCGACAGATTG 786
DB 121 GGTGTCCCGCGGATTTGGTCTAATCGACGAGCGTTGGCAATCGATTTGACTGCTCC 180
QY 787 GATCCATACCAAGAAGATGAACCAACCGTGGCGGCGAGCAATGCCCTCCGT 846
DB 181 GATGATATGATGTTGGAAGGATGAGTTGACCTGCGCGGGAGCAAAATGCTTCCAGG 240
QY 847 CTTTGAATTCACAGAAATTAACAATTCGCTAGCTAGCTCAATCCCAAGGCGACCGCG 906
DB 241 CTTCCGAAATTTTCAGAGAACTTCAAGTTCAAGACTACGCTCTCCGAAAGCA----- 295
QY 907 CCCCAGCCGCCAGAAAGGATGAAGGCGTTTATAGATGAACCTCAAGAGAGATTAG 966
DB 296 -----AAAGAGAGTGGGATGAAGGCTTTCGACAGATCTGAAGAAGATTCTCC 348
QY 967 ACTGTGGAGCATGTTTATTTGGCATGCTTTGTGATANTGGGGTGCGCTCCCGC 1026
DB 349 ACCGTGATATACATCTACCTGCGACGCGCTTTGGCGTACTGCGGAGGTCTTCTGCC 408
QY 1027 CAGGTGCGGCTGCTGCTAGCGACGTGATTCACACCACTGCTTCCACAGGGCTGAG 1086
DB 409 GGAGCTCTACTCTCCCGCTTCTACTATTTGTCGRCACAGAGTCTTCGCGGGGCTTAA 468
QY 1087 ATGACGATGAGAGATTGGCGGTGATGATGTTCTTCATAGGTCGGCGGTCCGCG 1146
DB 469 TTGACGATCAAGATCTCGCGCTTGTATAGATCATCGATACCGGAATCGGATTCTGTCG 528

QY 1147 CCGAGAGCGCTGAGGATGTACGAAGCACTTCATGCTCATTTGGAAAAAGTTGGGATC 1206
DB 529 CCGGACATGGCGAAACGAGTTTATCGAAGGTCTTCACTCATCTCATCAAAACGTCGAT 588
QY 1207 GACGGTTTAAAGTTAGCTTATCCACTTATGGAGATGTTGTGTGAACCTATGAGAG 1266
DB 589 AACGGGCTTAAAGTTAGCTTATCCACTTATGGAGATGTTGTGTGAACCTATGAGAG 648
QY 1267 AGAGTGGATTGGCAAGGATATACAAAGCAATGACCAAAATCAATTAATTAACATTT 1326
DB 649 AGAGTTGACTTGGCTAAAGCTTACTTCAAGCGTTAAGCGTCAGAGATTAAGCATTT 708
QY 1327 AAAGAAATGAGATCATTTGCAAGTATGCAACATTTGAACGACTTCACTTCTTGGCAG 1386
DB 709 GAGGCAACCGCTTATTCGACACATGAGACACTGTAATGACTTCAATGTTCTTGGAAC 768
QY 1387 GAAGTATCTCTTGGTCTGTTGGTATGACTTTTGGTGCACGGACCCCTGGGAT 1446
DB 769 GAAGCCATCTCTAGGCTGTGGGATGATGCTTGGTGCACGGATCCATCTGGGAC 828
QY 1447 CCAAGCGTACGTTTGGCTCCAAAGATGTCACATGGTTCATTTGTCACAGACACTTG 1506
DB 829 ATTAACGGCACGATTTGGCTGCAAGATGTCACATGGTTCACATGCTGCTACACAGTCT 888
QY 1507 TGGATGGGACATTCATCCACCTGACTGGGATATGTTCCATTCACACCACTTGTGCC 1566
DB 889 TGGATGGGAAATTTATCCAGCCTGATTTGGACATGTTTACGACACATCTTGTGCT 948
QY 1567 GCCTTCATGCTGCTGTCAGCATCTGTCGGGCGGATGATGTTATGATGTTGCTG 1626
DB 949 GAGTTCCATGCTCTTCACTGCTGCAATCTCCGTTGGGCCCATTTACATCAGCATTTGCTG 1008
QY 1627 GGAAGCATTAATCTTATCTTCTGAAAAAAACTAGTCTCTGATGATGATCTCTTCA 1686
DB 1009 GGCACAGACATTTGATCTCTTGAAGAGACCTGTTTGGCTGACGTTGATTTGAGG 1068
QY 1687 AGTGAATCTATGACTCCGACTCCGAGATTGTTTGTGAACACCTTTGCTTAATGA 1746
DB 1069 TGTGATCTATCTCTCTCCCACTGTCGACCTCTCTTTGMAACCTCTCTGATGAGC 1128
QY 1747 GAAATGATGCTTAAGATTGGAATCTCAACAGTCACTGAGTGAATGGTGATTCAC 1806
DB 1129 AAAACATGCTCAAGATTGTGAACCTTGAACTGAACTACATGGAATCATCGAGATTCAC 1188
QY 1807 TGCAGAGAGAGATGTTGCTGAGACACGCCGCAACCAATGCTTTCACAATCTCA 1866
DB 1189 TGTCAAGAGAGAGATGTTGACAGAAACTCGACGACCAATGCTTTCCCAATGCTT 1248
QY 1867 AAACGATGATCTCAAACTAACCCAAAAGACATGAGATGCGACAGTGAAGAAACCT 1926
DB 1249 AACAGCTTAAACCCCAACAAATCTTAATGAGCTTGAATGAACAGTGGGAACAAACCG 1308
QY 1927 ATCTATTGAAGGCGTTTAAACCTTGGCGCTTACTCTATCAAGCAAAACCTTATC 1986
DB 1309 ATCTCATTTGAAGCTTGAAGATTGCTTGTCTTGTCTCAATTCAGAAAGCTTTG 1368
QY 1987 CTCTCCAGCCCTCTCAAGATCTTGACATAGCTTGAACCATTCGATTCGATGAGCTATC 2046
DB 1369 TTGTCCGGGCAAAACGATGATCTCGAATCACAATGAGCCCTTCAAGTTCCGAGCTATC 1428
QY 2047 ACTGTTTCAACAGTACCAAACTCAACCAACTTCTTACACTTTGCCCAATTTGGGCTG 2106
DB 1429 ACTGTCACACAGTGTTCACCATTTGAGGGAGTTCGTTCAAGTTCTCCCAATCGGATG 1488
QY 2107 GTGACATGCTTAAACCTGTTGAGGCAATTCGTTGAGCAATGAGCAATGAGCAATG 2166
DB 1489 GTTAACATGCTTAAACCTGTTGAGGCAATTCGTTGAGCAATGAGCAATGAGCAATG 1542
QY 2167 TCAGTCGATGATGTTGCAAGAGTGTGATGATGATGATGATGATGATGATGATGATG 2226
DB 1543 TCCGTTGAGATGCGTGTGCTGAGGAAATTCAGGGTTTATGATCATCGAAGAAACCT 1602
QY 2227 AGGCTTGTCTGATGATGAGGAGATGTTGGGTTCAAGTATGATCAGACCAAAATGCTG 2286

Db 1603 GTGAGCTGCAGATGATGATGAAGATGTTGACCTTGGTAGC---AAGAGTCATAGTGTG 1659
QY 2287 GTGCTTCAGATGCCATGCGC 2306
Db 1660 ATGGTTCAAGTGCCTTGTC 1679

RESULT 15
AA220207
ID AA220207 standard; cDNA; 928 BP.
AC AA220207;
DT 17-JAN-2000 (first entry)
XX Soybean raffinose synthase cDNA.
DE Soybean raffinose synthase cDNA.
XX Raffinose synthase; soybean; transgenic plant; ss.
XX Glycine max.
OS Glycine max.
FH Key Location/Qualifiers
FT CDS 2..799
FT /tag= a
PN EP953643-A2.
PD 03-NOV-1999.
XX 27-APR-1999: 99EP-0107430.
XX 30-APR-1998: 98JP-0120550.
PR 30-APR-1998: 98JP-0120551.
PR 04-DEC-1998: 98JP-0345590.
PR 10-DEC-1998: 98JP-0351246.
XX (SUMO) SUMITOMO CHEM CO LTD.
XX Watanabe E, Oeda K;
PI MPI: 1999-593144/51.
DR P-PSDB; AAY32072.
XX New sense and antisense genes, useful for altering the level of
PT raffinose in food plants -
XX Claim 3; Page 20-22; 55pp: English.
XX This is the nucleotide sequence of a soybean cDNA clone coding
CC for raffinose synthase (see AAY32072), a protein which can bind a
CC D-galactosyl group through an alpha(1-6) bond to the hydroxy group
CC attached to the carbon atom at the 6-position of the D-glucose
CC residue in a sucrose molecule to form raffinose. The cDNA was
CC isolated from soybean cv. Williams 82 immature seed cDNA by PCR.
CC Probes or primers generated from plant raffinose synthase genes
CC (see AA220207-10) may be used to obtain other raffinose synthase genes
CC by labeled detection or amplification (claimed). These genes may
CC be used to control the levels of raffinose produced in plants.
CC Antisense genes can be used to knock out existing gene activity,
CC and sense genes to increase the level of gene activity. The
CC resulting transgenic plants may be used as a food source to alter
CC the growing conditions for gut enterobacteria, providing general
CC health advantages. The sense and antisense genes may also be used
CC in gene and phenotypic analysis of plants and for the selection of
CC plant clones with the desired characteristics with respect to
CC raffinose content.
XX
SQ Sequence 928 BP; 266 A; 171 C; 222 G; 269 T; 0 other;

Query Match 12.9%; Score 303.8; DB 20; Length 928;
Best Local Similarity 63.0%; Pred. No. 8; 9e-87;
Matches 486; Conservative 0; Mismatches 282; Indels 3; Gaps 1;

QY 1545 CCAATCCACCACCCCTTGTGCGCCTTTCATGCTGCCTCGAGCCATGCTGTGCGCC 1604
Db 1 CCAATCTGATCATGCTTGTGCGCAATTCACGCTGCTTACAGCCATTTCTGTGAGCC 60
QY 1605 GATCTATGTTAGTATGATTCGTGGAAAGCATTAATTGATCTTCTGAAAAAATAGTGTCT 1664
Db 61 AATTATATGTAGCGACTGTGTGGAAACACAACTTCAAGTTGCTTAAGAAGCTTGTTC 120
QY 1665 TCCCTGATGATGATCCTTTCGAAAGTGAATGATCTATGACATCCGACATCGGCGATTGTTT 1724
Db 121 ACCGTATGCGCTCCATTTTGTGGGTGCACATATGACATTCGCCACCGAGACATGCTTAAT 180
QY 1725 TGAAGACCCCTTGTGATTAATGAGAACTATGCTTAAGATTGGATATCTCAACAGTTTAC 1784
Db 181 TGTAGATTCCTTTACATGATGAGAAACAACTCTCAAAATTTGGAACTCATTAATATTTC 240
QY 1785 TGGAGTGAATTTGTGATTCACACTGCCAAGAGAGAGATGTCGTGAGACACGCGCAA 1844
Db 241 CGGGGTTTGGGTCTGTGTTCAATTTGCCAAGAGAGAGGTGTCCTGTACTAGGCGAAA 300
QY 1845 CCAATGCTTTTCAATACATACCAAAACGAGTGCATCCAAACTTAACCAAGACATAGA 1904
Db 301 CAAGAGTAGCTGTGACTATTCACACTCCGAGACTTGTGCAAGTCTTCMAAGACATTTGA 360
QY 1905 ATGGCACAGTGAGAAACCCCTATCTCTATTTGAAGGGGTTAAACCTTTGCGCTTACCT 1964
Db 361 ATGGGCGAAAGGAGACACCCACTTTGCAATCAAAAGGGGTGAGACTATTTGCTGTACAT 420
QY 1965 GTATCAAGCCMAAACTTATCTCTCCAAAGCCCTCTCAAGATCTTGACATAGCTTTGA 2024
Db 421 GTTTAAGAGACGACAGTTGAAGCTGCTGAAGGTACACAGAGAGTGTGAAGATTCTCTTGA 480
QY 2025 CCCATTGCAATTCGAGCTCATCTGTTTACACAGTACCAAACTCATCAAACTTCTCT 2084
Db 481 GCGTTTGTAGTTGTGACCTTTTGACCGTTTCTCCAGTGTGATCTTACCCGAAATCAAT 540
QY 2085 ACACTTTGCCCCCAATTTGGGCTGTGAACATGCTTAACACTAGTGAAGCCATCAATCTGT 2144
Db 541 CCAATTTGCCCAATTTGATTTGATTAACATGCTCAACTGTGGGGGCTCTATTATGTCAAT 600
QY 2145 GGACTTTGACGATGACCTTAAGCTCAAGTTCGAGATTGTTCAAAAGGCTGTGAGATGCG 2204
Db 601 GGAATTTGATCAACAGGAAATTTGGCGAGATTGGGGTGAAGAGACATGCGGAAATGAG 660
QY 2205 AGTATTTGCATCGAAAAAACCAAGGGCTTGTCTATTTGATGGGAGAGATGTTGCA 2264
Db 661 GGTATTTGCATCGAAGAGCCAGAGAGTGTCAAGATTGATGAGAAATCTGTGGAATTTGA 720
QY 2265 GTATGATCAAGACCAATGTTGTTGTTCAAGTGCATGCGCAATTGATTC 2315
Db 721 TTATGTT---GATAGAACGTGAGGCTCCAAAGTCTGTGTGCTTCTTC 768

Search completed: April 3, 2003, 16:22:10
Job time : 627 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 16:04:40 ; Search time 3761 Seconds
(without alignments)
10128.100 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407
Perfect score: 2352
Sequence: 1 ATGGCTCCTGAGTTTAA...CGTTATCGACTGTTT 2352

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rnd:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	416	17.7	668	12	BF635461 NF080A03D
2	402	17.1	646	9	A1727515 BNLGH1831
3	371.6	15.8	649	12	BF636478 NF089A04D
4	363.2	15.4	616	10	BE131081 L48-1031T
5	339.6	14.4	515	9	A1727581 BNLGH1847
6	335.4	14.3	682	12	BF634464 NF061C04D

7	322	13.7	559	13	BI974727	BI974727 sa172D10.
8	304.2	12.9	564	12	BE269676	BE269676 GA_EB000
9	301.4	12.8	560	14	BM887415	BM887415 sam38B08.
10	295.6	12.6	489	10	AV421962	AV421962 AV421962
11	293.4	12.5	513	12	BF636592	BF636592 NF090H02D
12	274	11.6	818	12	BG321136	BG321136 ZM04_05d1
13	268.2	11.4	528	14	BM886707	BM886707 sam29B05.
14	268	11.4	525	12	BF425566	BF425566 su44c11.Y
15	266.6	11.3	530	9	AT993310	AT993310 701496022
16	264.4	11.2	435	12	BF633327	BF633327 NF054D03D
17	264	11.2	486	10	AW944715	AW944715 00168_1ea
18	259.2	11.0	714	14	BQ969695	BQ969695 QHB39E15.
19	251.2	10.7	591	12	BG043590	BG043590 su40f05.Y
20	248.8	10.6	695	12	BF633311	BF633311 NF0596C08D
21	237.8	10.1	672	12	BG454761	BG454761 NF104H06L
22	236.8	10.1	558	12	BG600526	BG600526 EST505421
23	236.6	10.1	435	10	BE248707	BE248707 NF010D10D
24	233.4	9.9	801	17	BH483423	BH483423 BOC2056TR
25	211.2	9.0	518	10	AW649251	AW649251 EST327705
26	210	8.9	597	12	BG450458	BG450458 NF014C06D
27	203.2	8.6	832	14	BQ508902	BQ508902 EST616317
28	195.8	8.3	541	14	BM887020	BM887020 sam33C03.
29	190.8	8.1	714	14	C72350	C72350 C72350 RICE
30	188.6	8.0	592	12	BF632552	BF632552 NF027E01D
31	187.4	8.0	540	12	BG455277	BG455277 NF046E04P
32	185.6	7.9	375	10	BE187139	BE187139 NXNV_159
33	183.6	7.8	556	12	BF632381	BF632381 NF027B09D
34	181.6	7.7	557	12	BF480406	BF480406 LO-2338T3
35	181	7.7	496	9	AU083889	AU083889 AU083889
36	177.6	7.6	635	10	AV538951	AV538951 AV538951
37	174.2	7.4	411	14	T75887	T75887 10665_Lambd
38	170.4	7.2	610	12	BG648423	BG648423 EST510042
39	168.4	7.2	322	14	DA0721	DA0721 RICE5852A_R
40	166.4	7.1	671	12	BG452557	BG452557 NF107B10L
41	166	7.1	307	12	BF516701	BF516701 NX51_002
42	163.8	7.0	360	9	A1442026	A1442026 sa66f09.Y
43	162	6.9	446	10	BE346804	BE346804 sp31d06.Y
44	160.2	6.8	700	13	BI406578	BI406578 171E01 Ma
45	153.2	6.5	830	17	BH483417	BH483417 BOC2056TR

ALIGNMENTS

RESULT 1
LOCUS NF080A03D1F1020 Drought Medicago truncatula cDNA clone NF080A03DT
DEFINITION 5', mRNA sequence.
ACCESSION BF635461
VERSION BF635461.1 GI:11899619
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicotyledons;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifoliaceae;
Medicago.
1 (bases 1 to 668)
Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Imman,J.T., Weller,J.W. and May,G.D.
Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
Unpublished (2000)
Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert length: 668 Std Error: 0.00
Plate: 080 row: A column: 03

FEATURES	Seq primer: TCACACAGGAACACACTATGAC.
source	Location/Qualifiers
	1. 668
	/organism="Medicago truncatula"
	/db_xref="taxon:3880"
	/clone_id="NF080A03Df"
	/clone_lib="Drought"
	/tissue_type="Plantlets"
	/dev_stage="Pooled timepoints"
	/note="Vector: Lambda Zap; Contains a mixture of entire plantlets harvested in a series of days-post-watering timepoints."
BASE COUNT	181 a 125 c 148 g 214 t
ORIGIN	
Query Match	17.7%
Best Local Similarity	77.9%
Matches 514:	Conservative 0; No.M.1e-11; Indels 1; Gaps 1

[illegible]

ORGANISM	
	Gossypium hirsutum Eukaryota: Viridiplantae: Streptophyta: Embryophyta; Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots; Rosidae: eurosids II: Malvales: Malvaceae: Gossypium.
REFERENCE	1 (bases 1 to 646)
AUTHORS	Blewitt,M., Matz,E.C., Davy,D.F. and Burr,B.
TITLE	ESTs from developing cotton fiber
JOURNAL	Unpublished (1999)
COMMENT	Contact: Ben Burr Biology Department Brookhaven National Laboratory Upton, NY 11973, USA Tel: 516-344-3396 Fax: 516-344-3407 Email: burr@bnlxi.bnl.gov Seq primer: 13 primer.
FEATURES	Location/Qualifiers
source	1..646

FEATURES	Seq primer: T3 Primer.	Location/Qualifiers
Source	1. 646	
	"/organism="Gossypium hirsutum" "/cultivar="Acala Maxxa" "/db_xref="taxon:3635" "/clone.lib="Six-day Cotton fiber" "/tissue_type="Immature fiber" "/dev_stage="Six days post anthesis" "/lab_host="XL1-Blue" "/note="Vector: pBluescript II KS+"	
BASE COUNT	162 a	139 c 149 g 193 t 3 others
ORIGIN		
Query Match	17.1%;	Score 402; DB 9; Length 646;
Best Local Similarity	76.3%;	Pred. No. 1.4e-107;
Matches 492; Conservative	0;	Mismatches 153; Indels 0; Gaps 0;
QY 1083	GCAGATGACGATGAGAGATTGGCGGTGATTAAGATTGTTTCATTAAGCTGGCGCTGCT	1142
Db 1	GAGAAACGATGAGAGATCTTCGGCTCATTAAGATTGCATATTGGTATGCGATTAGT	60
QY 1143	CCCGCCGGAAGAGCGTAGAGATGTACGAAGACTTCATGCTCATATTGGAAAAAGTTGG	1202
Db 61	GCGCGCGGAGATGGCTGATCACTTATGSAAGAAATCTTCTCATTTGGAAAAATGTTGG	120
QY 1203	GATCGACGGTGTAAAGATTGACGTTATCCACTATTGGAGATGTTGTGGAAGACTATGG	1262
Db 121	AATTACGGAGATTAAAGTTGACGTATTTCATCTGTTGGAAAAAGCTGTGGAAAACTATGG	180
QY 1263	AGGGAGAGTGGATTGGCAAGCAATATTACAAGCAAGTACCACCAATCAATAATTAAACA	1322
Db 181	TGGAAAGTGTATCTTGGCAAAAGCTTATTATTAAAGCTTTAAACAGATTCAAGTTAAAGA	240
QY 1323	TTTTTAAGGAATGAGATCATATGCAAGTATGAGACATTGTAAACGACTTCATGTCCTTGG	1382
Db 241	TTTTCAAGGCATGTGCTTATTATGCCAGTATGGAACACATCCAAACGATTCATGTTCTTGG	300
QY 1383	CACGGAAGCTATCTCTCTTGGTCGTGTTGGTGATGACTTTTGGTGACAGGACCCCTTGG	1442
Db 301	AACAGAAGCATTTTGTCTTGGTCGTGTCGGAGCATTTTTTGGTCACATGCATCCATCAGG	360
QY 1443	TGATCCAAACGTTAGCTTTTGGCTCCAAAGATGTCAACATGTTCAATGTGCCAACGACAG	1502
Db 361	TGACCTTATGGGACATTTTGGCTCCAAAGTTGTACACATGGGCACTGTGTTACACAG	420
QY 1503	CTTGTGATGGGGAATCATCATCCACCTGACGTAGATATGTTCCATCCACACCCTTGG	1562
Db 421	TTTATATGATGGGCAACTTATATCCACCCGTGATTTGGGACATGTTCCAGTCTCTCCACCCCTGG	480
QY 1563	TGCCGCTTCATGCTGCGCTCTCGAAGCATCTCTGTGGGCCGATCTATGTTAGTGAATTC	1622
Db 481	TGCTAGATTCCACGCGCTTCAAGGGCCATCTCCGGTGGCCCAATTTACATCANGACAGC	540
QY 1623	TGTGGGAAGCATTAATCTTGTATGAAAAAACTAGTGCCTTCTGATGATGATCATCT	1682
Db 541	TGTTGGCAACCAACATTTGCTCTCTCCAAAGCCCTCGTNTTACCAGGATGTTGCATCTCT	600

JOURNAL Unpublished (2000)
COMMENT Contact: May GD

BASE COUNT	196 a	127 c	143 g	215 t	1 others
ORIGIN					

Query Match	14.38;	Score 335.4;	DB 12;	Length 682;
Best Local Similarity	69.38;	Pred. No. 8.5e-88;		
Matches 470;	Conservative 0;	Mismatches 207;	Indels 1;	Gaps 1

Oy	1570	TTTCATCTGCTCTGAGACATCTCTGGGGCCGATATATGTAAGATTTGTGGGA	1629
Db	2	TTTATGACACTCTCTAGAGCCATATCTGGTGGTCAATTTACATATAGTGACACTGTGGA	61
Oy	1630	AAGCATAACTTGTGATCTCTGAAAAAACAAGTACGTCTCTGATGATCGATCCCTCGAAGT	1689
Db	62	AATCAACAATTTTGGACCTTCTCAGAAACACTAGCTTTGCCGATGGTAACCATCTTATGATGT	121
Oy	1690	GAGTACATACACTCCCGACTCCCGCATGTGTGTTTGAAGAACCCCTTGTCATTAATGAGAA	1749
Db	122	GAACCATATGCTCTCCCACTAATGAGACTGTCTTTTGTGATCTCTTTCACATGATGGCAAA	181
Oy	1750	ACTATGCTTAAATTTTGGATCTCAACAAGTTCACGTGAGTATTTGGTCAATTCACATGCG	1809
Db	182	ACTATGCTCAAAATTTTGGAACCTCATATAGTACACTGGAGTTCTTGGTGTGTTTAACGCG	241
Oy	1810	CAAGGAGAGAGATGTGTCTGTGAGACACGCCCAACCAATGCTTTTCACATATCTCAAAA	1869
Db	242	CAAGGAGAGAGATGTGTCTCCGAGATTAAGATCAACAAATGTGCTCTGAGTTTTCACAT	301
Oy	1870	CGAGGACATCCAAAACCTTAACCCAAAAGACATGAAATGGCACAGTGGAGAAACCCCTTC	1929
Db	302	TGCGTATCACTAAGATCAACATCAAGACATTTGAATGGAACGTGGAGAGATTCCAATT	361
Oy	1930	TCTATTTGAAGCCGTTAAAAACCTTTGGCGTTTACCCTTATCAAGCCAAAAAATTATCCTC	1989
Db	362	CCCATTTGAAGAGTTTCAAGTTTTCCTGTTGATTTTCAAGCAAGCCAAAGCTGAACTT	421
Oy	1990	TCCAAAGCCCTCTCAAGATCTTTGACATAGCTCTTGACCCATTTGGAATTTGAGCTCATCT	2049
Db	422	TTTCTCTCCATCTGAAATTTGAAGAAATCTCTTTGGACCACTTCAACCTTGAGCTTTAACT	481
Oy	2050	GTTTTCACCAAGTGAACCAATCATCAACAACTTCTCTACATTTTGGCCCAATTTGGCGTGGT	2109
Db	482	GTTTCTCTCAAGTCACTTATTTTCTCTAAAAAGTCCCTTAATGTTTGCCTCTATTTGGTTGGTT	541
Oy	2110	AACATGCTTAACTACTAGTGAAGCCATTCATCTGTGGACTATGACAGATGACCTTAAGTCA	2169
Db	542	AACATGCTAAACAATGCTGGGGCAATTCAGATCTTTGGAATATCTTGAGGCTCAAGATTTG	601
Oy	2170	GTCGAGATGGTGTCAAAAGGCTGGTGGAGATG-CGAGTATTTGCAATCGAAAAAACGAG	2228
Db	602	GTCGAGTGGAAATTAAGAGTGGCTGGTGGAGATGAGGGGTCTATGCTTCAMAGAAAAACGAG	661

Qy	2229	GGCTTGTGCGTATTGATGG	2246
Db	662	GGCTTGTGGAATTGATGG	679

RESULT 7	BI974727	559 bp	mRNA	linear	EST 30-NOV-2001
LOCUS	BI974727				
DEFINITION	sa172b10.y1 Gm-c1068 glycine max cDNA clone				
FEATURES	ID: Gm-c1068-3932 5' similar to TR:Q92762 Q92762 RAFFINOSE SYNTHASE				
ORIGIN	11 mRNA sequence.				

ACCESSION	BI974727	GI:16349132
VERSION	BI974727.1	
KEYWORDS	EST.	
SOURCE	soybean.	
ORGANISM	Glycine max	

REFERENCE
1 (bases 1 to 559)
Shoemaker, R., Keim, P., Vodkin, L., Eprelding, J., Corryell, V., Khanna

TITLE	JOURNAL	COMMENT
Public Soybean EST Project	Unpublished (1999)	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University, School of Medicine 444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntsville, AL 35801 For further information call: (800)-533-4363 or contact via email: ccu@resgen.com High quality sequence stop: 473.

```

FEATURES
source
Location/Qualifiers
1..559
/organism="Glycyne max"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1068-3932"
/clone_l1b="Gm-c1068"
/tissue_type="leaf, drought stressed, 1 month old plants, greenhouse grown"
/lab_host="DH10B"
/note="Vector: pBluescript II SK+ Site.1: EcoRI; Site.2: XhoI. The cDNA library was constructed from mRNA isolated from drought stressed leaf tissue of the cultivar Williams 82. The month old greenhouse grown plants were deprived of water for 3 days prior to harvesting the stressed leaf tissue. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dI) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker."

```

BASE COUNT	147 a	135 c	129 g	148 t
ORIGIN				
Query Match		13.7%	Score 322;	DB 13;
Best Local Similarity		73.8%;	Pred. No. 7.3e-84;	
Matches 409; Conservative		0;	Mismatches 145;	Indels 0; Gaps 0

QY 1459 TTTTGGCTCCAGGATGTCAATGCTTCATTCTGCCAACGACAGCTTGTGGATGGGGAAC 1518
||||||| ||||| ||||||||| ||||||| || ||||||||| ||||| ||
Db 2 TTTTGGCTACAGGCTGTCAATGCTTCATTGTGCATACACAGCTTGTGGATGGGCAAT 61

```

OY 1519 TTCATCCACCTGACTGGATATGTTCCAATCCACCCACCTGTGGCTTCATGCT 1578
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 62 TTCATCCACCAAGATTGGGACATGTTCCAATCAGTACATCCTGTGCTTCATGCT 121
OY 1579 GCCTTCGAGACCATCTCTGTGTGGCCGATCTATGTTAGTATTTCTGTGGAAGCATAC 1638
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 122 GCCTCAAGAGCCTATATCTGTGTGGCCCATTTACATCAGTACACAGTGTGGAGCACCAAC 181
OY 1639 TTGTGATCTTGTAAAAAATAGTCTTCCGATGATGCATCCTGTGAAGTGTACTAT 1698
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 182 TTGTGACTGCTTATAGACCTTTGGCTTGCAGATGGGTCCATCTCATGATGTAGCACTAT 241
OY 1699 GCATCTCCGACTCGCATGCTTTGTTTGAAGACCTTTGCTATATAGAGAATACTATCTT 1758
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 242 GCACCTCCCAACGAGGAGCTGTCTTGTGTGACCCCTCTCATATGTGGCAAAACATATGCTC 301
OY 1759 AAGATTGGAATCTCAACAAGTTCACTGAGTGTATGTGTGCAATTCAGTCCAGAGAGA 1818
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 302 AAGATATGGAACCTCAACAAGTACACTGGATGTTCTTGGGGTGTTTAACTGCGGAGAGA 361
OY 1819 GGATGCTGTGTGAGACAGCGCCGACCAATGCTTTCAACAATACCAAAAGAGAGA 1878
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 362 GGTGTGCTTCCGTAGATTAGTTCACCAAAATGTGCTGTGAGTTTCTCATAGGATATCA 421
OY 1879 TCCAAAACCTAACCAAAAGACATAGATGACAGAGTGGAGAAAACCTATCTTATTTGAA 1938
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 422 ACCAAGACCAATATCAAAAGACATTTGAATGAGATAGTGAAGATTCATTTCCATTTGAA 481
OY 1939 GGCCTTAAACCTTTGGCTTTACCTCTATCAAGCCAAAACCTTATCTCTTCAAGCCC 1998
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 482 GGGGTGACACTTTTGTGCTTGTATTTACGACCAAGCAAGAACTCATCTCAGCAGCA 541
OY 1999 TCTCAAGATCTTGA 2012
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 542 TCTGATGACAGTGA 555

```

RESULT 8
LOCUS BF269676 564 bp mRNA linear EST 07-MAR-2001
DEFINITION GA_EB0005F08f Gossypium arboreum 7-10 dpa fiber library Gossypium
ACCESSION BF269676
VERSION BF269676.1 GI:11200671
KEYWORDS EST.
SOURCE Gossypium arboreum.
ORGANISM Gossypium arboreum.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Malvales; Malvaceae; Gossypium.
REFERENCE 1 (bases 1 to 564)
AUTHORS Wing,R.A., Frisch,D., Yu,Y., Main,D., Rambo,T., Simmons,J., Henry
 D., Wood,T.C., Leslie,A. and Wilkins,T.A.
 An integrated analysis of the genetics, development, and evolution
 of the cotton fiber
 Unpublished (2000)
JOURNAL Contact: Wing RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson, SC 29634, USA
 Tel: 864 656 7288
 Fax: 864 656 4293
 Email: rwing@clemson.edu
FEATURES Seq primer: TAATACGACCTACATAGG
 High quality sequence stop: 563.
 Location/Qualifiers
 1..564
 /organism="Gossypium arboreum"
 /strain="AKA"
 /cultivar="8400"
 /db_xref="taxon:29729"
 /clone="GA_EB0005F08f"

```

/clone.lib="Gossypium arboreum 7-10 dpa fiber library"
/feature_type="fibers isolated from bolls harvested 7-10
  dpa"
/lab_host="E. coli"
/Note="Vector: pBK-CMV; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT 149 a 143 c 124 g 145 t 3 others
ORIGIN
Query Match 12.9% Score 304.2; DB 12; Length 564;
Best Local Similarity 72.2%; Pred. No. 1,4e-78;
Matches 407; Conservative 0; Mismatches 156; Indels 1; Gaps 1;

```

QUERY 1453 GGTACGTTTGGCTCCAA-GGATGTCACATGATTCATTTGGCAAGACAGCTTGGAT 1511
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 GGGACATTTTGGCTCAAGGTTGTACATGATGACACTGTGCTTAACACAGTTTATGAT 60
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1512 GGGGAACCTTCATCCACCTGACTGGATATGTTCCAAATCCACCCCTTGTGCCCTT 1571
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 GGGCAACTTATATCCACCCGATTTGACATGTTCCAGTCTCCACCCCTTGTGCTGATT 120
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1572 CCATGCTGCTCGAGCCATCTCTGTGGCCGATCTATGTTAGTATTTCTGTGGAAA 1631
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 121 CCAGCTGCTTCAAGGGCATCTCGGTGGCCCAATTTACATCAGGACACTGTGGCAA 180
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1632 GCATTAATTTGATCTTGTGAAAAAACTAGTGTCTCTGATGATCGATCTTCAAGTGA 1691
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 181 CCACAACTTTGCTCTCTCAAAAGCCTGTTTACCGATGTTGATCTCTCCGCTGCCA 240
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1692 GTTATGACATCTCCGACATCGGATTTGTTTGAAGACCTTTGCATATGAGAAAC 1751
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 241 ATACTAGCTCTTCCAAACGAGGATTCCTATTGTGAGACCCTTCATGATGGCAAAAC 300
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1752 TATGCTTAATTTGGAATCTCAACAGTTCACTGGAGATGTTGGTCAATTCAGTCA 1811
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 301 CATGCTCAAAATCTGAACTTGAATAGTATCTGTGTGATTTGGCGATTTTAACTGCCA 360
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1812 AGGAGAGAGATGATGTCTGAGACAGCCGCAACCAATCTTTTCAATTAAGTCAAAAG 1871
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 361 AGGAGCGGATGATGTGCGGAGAAATGAAAGAAACCAATCTTTTCCGATTTTCCGATAC 420
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1872 AGTGACATCCAAACTTAACCCCAAGACATAGATGACAGTGGAGAAACCTTATCTC 1931
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 421 GGTAAAGGCGGAGATGAATCAAGAAACATGTAGTGGAACAGTGGAAAGACCAATTTTC 480
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1932 TATGGAAGCGCTTAAACCTTTGGCTTACCTATCAAGCAAAAACCTATCTCTC 1991
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 481 CATTGAAGATGTCCAAGTATTTGGCCATGTATTTTCTCAGTCCAGAAAGATAGTCTCTC 540
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1992 CAAGCCCTCTCAAGATCTTGACAT 2015
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 541 ANAACCGCTGAAGAACATGCAAT 564
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
LOCUS BM887415 560 bp mRNA linear EST 08-MAR-2002
DEFINITION sam38D08.v1 Gm-cl068 glycine max cDNA clone SOYBEAN CLONE ID:
 Gm-cl068-6976 5' similar to TR:Q9ZT62 Q9ZT62 RAFFINOSE SYNTHASE ;,
 mRNA sequence.
ACCESSION BM887415
VERSION BM887415.1 GI:19271159
KEYWORDS EST.
SOURCE soybean.
ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
REFERENCE 1 (bases 1 to 560)
AUTHORS Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna
 A., Bolla,B., Maria,M., Hillier,L., Kuabla,T., Martin,D., Beck,C.,
 Wylie,T., Underwood,K., Stepien,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk


```

Db 244 CCTGAGATGTGACACTATGACACTCCCACTAAGGACTGTCTCTTGATGATCTCTCA 303
Qy 1740 TAATGAGAACTATGCTTAAGATTTGGATCTCAACAGCTTACAGGATGTGCTG 1799
Db 304 TATGAGCAAAACAAATGCTCAAGATATGAACTCAACAGTACAGTGGATTTGGGT 363
Qy 1800 ATTCACTGCCAAGAGAGATGTGTCTGAGACACGCCCAACCAATGCTTTTCA 1859
Db 364 ATTTAAGTCCCAAGAGAGATGTGTCTGAGATCAATCCAAACAAAGTGTCTTGA 423
Qy 1860 AATCTCAAAACAGATGATCCAAATCAACCAAAATAGATGAGACAGTGA 1919
Db 424 GTTTTCTCATGTGATCAACAGATGATCCAAACATTTGAGTGAAGATGTGA 483
Qy 1920 AATCC 1925
Db 484 GAACC 489

RESULT 11
BF636592
LOCUS BF636592 513 bp mRNA linear EST 19-DEC-2000
DEFINITION NF090H02D1F1027 Drought Medicago truncatula cDNA clone NF090H02D1
5', mRNA sequence.
ACCESSION BF636592
VERSION BF636592.1 GI:11900750
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 513)
AUTHORS Torres-Jerez,I., Scott,A.D., Harris,A.R., Gonzales,R.A., Bell,C.J.,
Flores,H.R., Imman,J.T., Weiler,J.W. and May,G.D.
TITLE Expressed Sequence Tags from the Samuel Roberts Noble Foundation
Medicago truncatula drought library
JOURNAL Unpublished (2000)
COMMENT Contact: May GD
Plant Biology Division
The Samuel Roberts Noble Foundation
2510 Sam Noble Parkway, Ardmore, OK 73402, USA
Tel: 580 221 7391
Fax: 580 221 7380
Email: gdmay@noble.org
Insert length: 513 Std Error: 0.00
Plate: 090 row: H column: 02
Seq primer: TCACACAGAAACAGCTATGAC.
FEATURES
source
1..513
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone="NF090H02D1"
/clone_lib="Drought"
/tissue_type="Plantlets"
/dev_stage="Pooled timepoints"
/notes="Vector: Lambda Zap; Contains a mixture of entire
plantlets harvested in a series of days-post-watering
timepoints."
timepoints.

BASE COUNT 135 a 72 c 138 g 167 t 1 others
ORIGIN
Query Match 12.5% Score 293.4; DB 12; Length 513;
Best Local Similarity 73.2%; Pred. No. 2,1e-75;
Matches 375; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy 1000 TGTGATATTGGGAGCTTCGCCGACGAGTGCTGCTGAGCAGCTGTGATT 1059
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2 TGTGATTATTGGGAGGATTAAGACTAATGTTGAGGGCTTACCTGACCTATTGTGGAG 61

Qy 1060 CAGCAGTCTTTCACCAAGGCTGCAGATGACGATGGAGATTGGCGTGCATTAAGATT 1119

```

```

Db 62 AAGCCTAATTAACCGTGGATTTGAAACTACTATGAGAGATTGGCTGTATAAGATT 121
Qy 1120 GTTCTCATTAAGTGGCGGTGTCGCCGAGAAAGCGTAGAGATATACGAGACTT 1179
Db 122 GTTAATAATGATGTGGGTAGTCTCTCTCATTTGGTGAATCAAAATGATGAAGTATT 181
Qy 1180 CATGCTCATTTGGAAAAAGTTGGATGACGGTGTAAAGATTGACCTTTCACCTATTG 1239
Db 182 CACTCTCTTTGGAAAGAGCTGGAATTGATGGGCTCAAGGTTGATTCATTCATTGGCTA 241
Qy 1240 GAGATCTTGTGAAGACTATGAGAGGAGAGATTTTGGCAAAAGCATATTACAAGCA 1299
Db 242 GAGATGTGGTGAGAGATGATGTTGAGAGAGATTTGATTTGGCAAAACATATTACAAGCT 301
Qy 1300 ATGACCAATTCATTAATTAACATTTTAAAGAAATGAGTCAATTCAGTATGAGACAT 1359
Db 302 CTCTCACTTCAGTGAATAAACATTTTCATGTAATGATGATCTAGCATGAGACAT 361
Qy 1360 TGTAAAGCTTCAATGTTCTTGGCAGGAGAGCTATCTCTTGTGCTGTGGTGTATGAC 1419
Db 362 TCGAATGATTTTCATGTTGCTTGGAACTGAAGCCATTCCTTGTGTGTGTGATGAT 421
Qy 1420 TTTTGTGACAGGACCCCTCTGTGTATCCAAACGTTAGCTTTTGGCTCAAGATGTAC 1479
Db 422 TTCTGTGACAGTCAACCATATGATGATCCAAATGTGATATTGGCTACAGATGTAC 481
Qy 1480 ATGGTTCATTTGTCACAGACAGCAAGCTTGTGAT 1511
Db 482 ATGTCATTTGTCATATCAATATGATGAT 513

RESULT 12
BG321136
LOCUS BG321136 818 bp mRNA linear EST 27-FEB-2001
DEFINITION Zm04_05d11.R Zm04_AAFRC_ECORC_cold_stressed_maize_seedlings Zea mays
CDNA clone Zm04_05d11, mRNA sequence.
ACCESSION BG321136
VERSION BG321136.1 GI:13150814
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 818)
AUTHORS Singh,J.A., Makul,K., Couroux,P., De Moors,A., Harris,L.J., Hattori
J.I., Ouellet,T., Robert,L.S., Sprott,D. and Tinker,N.A.
TITLE Expressed Sequence Tags from Cold-Stressed Maize Seedlings
JOURNAL Unpublished (2001)
COMMENT Contact: Singh,J.A.
Eastern Cereal and Oilseed Research Centre
Agriculture and Agri-Food Canada
KW Nealey Bldg., Central Experimental Farm, Ottawa, Ontario, K1A
0C6, Canada
Tel: (613) 759-1662
Fax: (613) 759-1701
Email: singhja@em.agr.ca.
FEATURES
source
1..818
/organism="Zea mays"
/cultivar="CO328"
/db_xref="taxon:4577"
/clone="Zm04_05d11"
/clone_lib="Zm04_AAFRC_ECORC_cold_stressed_maize_seedlings"
/tissue_type="leaf, crown"
/notes="Vector: Bluescript SK-/XhoI-EcORI. Site_1: Eco RI;
Site_2: Xho I; Lower temperature 50 C / hour from 22 to
120C; bring to 50 in 1 hour from 120C. Leave at 50C 2 days
, photoperiod 16 hours. Light intensity was 125 uE-1.
library prepared by in vivo mass excision from amplified
library."

BASE COUNT 112 a 300 c 230 g 129 t 47 others

```


Accession	Sequence	Position
Oy	1786 GGAGTATGTGGCATTTAACTCAACGCCAAGAGAGATGGTGTGCGAGACAGCGGCAC	1845
Oy	1786 GGAGTATGTGGCATTTAACTCAACGCCAAGAGAGATGGTGTGCGAGACAGCGGCAC	1845
Db	241 GGAGTCTTGGGGTGTTTAACTGACGCCAGSAGSAGGTGGTGGCGAGATTAGGTCCAC	300
Oy	1846 CATCTCTTTTCAACATFACTCAAAACGAGTGACATCCAAATCTAACCCAAAGACATAGAA	1905
Db	301 AAATGTGCTGCGAGATTTTCTCTATGGGGTATCAACCAACCAATATCAAAACATTTGAA	360
Oy	1906 TGGCAGTGGAGAAACCCATCTCTATTTGAAGCGGTAAACCTTTGCGCTTAACTGC	1965
Db	361 TGGATTAATGGGAAAGATTCATTTTCCATTTGAAGGGGTCCAACTTTTGGCTTCGTAATTTG	420
Oy	1966 TATCAAGCCAAAACCTTATCTCTTCCACAGCCCTCTCAAGA--TCATTGACATAGCTTT	2022
Db	421 AGCCACAGCCAGAAACCTATCTCTACAGACCATCTGATGACAGTAGAGAGATTCCCTTG	480
Oy	2023 GACCATTGGAATGGAGCTCATCTACTGTTTACACAGTAGC 2063	
Db	481 GAGCCATTGAATTTGGAGCTTATTAACAGATTTTCCCTGTGAC 521	

RESULT	14
BFA25566	
LOCUS	BFA25566
DEFINITION	BFA25566 525 bp mRNA linear EST 06-DEC-2001
ACCESSION	SU44C1.Y1 Gm-cl068 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
VERSION	Gm-cl068-2061 5' smaller to TR:Q9ZT65 Q9ZT62 RAFFINOSE SYNTHASE ;
KEYWORDS	mRNA sequence.
SOURCE	BFA25566 BFA25566.1 GI:11413555
ORGANISM	EST.
	soybean.
	Glycine max

REFERENCE
AUTHORS
1 (bases 1 to 525)
Snoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna

TITLE	Public Soybean EST Project
JOURNAL	Unpublished (1999)
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel.: 314 286 1800

```
FEATURES      Location/Qualifiers
source        1. 525
```

Note: Vector: plbluecript II SK⁺, Site: 1: EcoRI, Site 2: XhoI; The cDNA library was constructed from mRNA isolated from drought stressed leaf tissue of the cultivar Williams 82. The month old greenhouse grown plants were deprived of water for 3 days prior to harvesting the stressed leaf tissue. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI

restriction site. EcoRI adaptors were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pluscript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in the laboratory of Dr. Randy Shoemaker.

[illegible]

RESULT	15
LOCUS	A1993310
DEFINITION	A1993310 530 bp mRNA linear EST 08-SEP-1999 701496022 A. thaliana, Ohio State clone set Arabidopsis thaliana
ACCESSION	CDNA clone 701496022, mRNA sequence.
VERSION	A1993310
KEYWORDS	A1993310.1 GI:5840215
SOURCE	EST.
ORGANISM	thale cress. Arabidopsis thaliana
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae: eurosids II; Brassicales; Brassicaceae; Arabidopsids. 1 (bases 1 to 530)
AUTHORS	Chen,J., Momiyama,M., Chan,E., Mooney,M., Carroon,B., Gilliland,D., Wang,X., Hillman,J., Guegler,K., Kim,C., Doyle,M., Brzoska,P., Gorgone,G., Burns,D., Griffin,J., Mounoutoua,M., Nguyen,D., Tan,R., Rose,M., Warren,B., Ton,B., Kastury,K., Borillo,C., Carpio,T., Policky,J., Suzuki,G., Argentine,C., Shah,S., Nobriza,A., Murry,L., Turner,C., Krikorian,S., Elder,L. and Hanson,D.
TITLE	Arabidopsis thaliana Gene Expression Microarray
JOURNAL	Unpublished (1999)

COMMENT

Contact: David Smoller, Ph.D.
Genome Systems, Inc., a wholly owned subsidiary of Incyte
Pharmaceuticals, Inc.
4633 World Parkway Circle, St. Louis, MO 63134, USA
Tel: 877-577-2733
Fax: 314-427-3324
Email: service@genomesystems.com

FEATURES

source

Location/Qualifiers
1. .530
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
/clone_lib="A. thaliana, Ohio State clone set"
/note="CDNA library was made from selected clones from the
Arabidopsis thaliana Ohio State clone set."

BASE COUNT

132 a 115 c 128 g 155 t

ORIGIN

Query Match

11.3%; Score 266.6; DB 9; Length 530;

Best Local Similarity 69.0%; Pred. No. 1.9e-67;

Matches 365; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

```
OY 1613 TTACGATTCCTGTGGGAAGCATATCTTGATCTTGAAAAACTAGTCTCTCGATG 1672
DB 2 TCAGGATTTGTGGGCAAGCATGATTTGATCTCTGGAAGCGCTGTTGGCCACG 61
OY 1673 GATCGATCCTTGCAGTGAATGATGCACTCCGCACTGCGATTTGTTGGAAGAC 1732
DB 62 GTTCGATTTTGAGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 121
OY 1733 CTTTCATTAATGAGAACTATGCTTAAGATTGGAATCTCAACAAGTTCACTGGAGTGA 1792
DB 122 CTCCTCATGATGGCAAAACCATCTCAAGATTGGAACCTGAACAAGTACACTGGAGTTA 181
OY 1793 TTGGTGCAATTCACATGCCAAGAGAGATGCTGTCGAGACAGCCGCAACCAATGCT 1852
DB 182 TTGGAGCATTCACATGCTCAAGAGAGATGCTGTCGAGAGAAACAGACGTAACCAATGTT 241
OY 1853 TTTCACATTAATCAAAAGAGATGATCCAAACTAACCCAAAGACATAGAAATGGACA 1912
DB 242 TCTCTGATTCCTCAACACGTTAACCGCCACCAACAGCCCTTAAGACGTTGAATGAAACA 301
OY 1913 GTGAGAAAAACCTATCTCTATTTGAAGCGGTTAAAACTTTGCGCTTACTCTATCAAG 1972
DB 302 GTGGAAGCAGCCCAATCTTCATTCGAAAGCTGGAAGAGTTGCTTGTCTGTCATAT 361
OY 1973 CCAAAAAAATCTCTCTCTCCCAAGCCCTCTCAAGATCTTGACATAGCTCTTGACCATTCG 2032
DB 362 CCAAGAGAGCTTTGTTGCTGTGAGCTAAACGATGATCTTGAGCTTGGAGCCCTTCA 421
OY 2033 AATTGAGCTCATCACTGTTTACCAAGTGACCAAACTCATCCAAACTTCTTACACTTTG 2092
DB 422 AGTTGAGCTGATCACTGCTCTCTCTCTGTTGATGACATGAGGTAATTCATCCGTTTG 481
OY 2093 CCCCAATTGGGCTGCTGAACATGCTTAACACTAGTGAGAGCCATCCAAATC 2141
DB 482 CTCGATTTGAGCTGGTAAACATGCTAAACACAAAGCGGTGATCCGGTC 530
```

Search completed: April 3, 2003, 19:22:25
Job time : 3798 secs



GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 16:07:45 ; Search time 100 Seconds
(without alignments)
7213.037 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407

Perfect score: 2352

Sequence: 1 ATGGCTCCTAGATTAA...CGGTATGAGTACTGTTT 2352

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq.*

2: /cgn2_6/ptodata/1/lna/5B.COMB.seq.*

3: /cgn2_6/ptodata/1/lna/6A.COMB.seq.*

4: /cgn2_6/ptodata/1/lna/6B.COMB.seq.*

5: /cgn2_6/ptodata/1/lna/6C.COMB.seq.*

6: /cgn2_6/ptodata/1/lna/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2352	100.0	2517	4	US-08-846-234-4
2	976.4	41.5	2498	4	US-09-561-825-23
3	47.4	2.0	7218	1	US-08-232-463-14
4	38	1.6	911	4	US-09-457-046B-19
5	38	1.6	911	4	US-09-457-046B-21
6	36.8	1.6	19736	4	US-09-740-035-3
7	36	1.5	2824	2	US-09-010-928B-3
8	36	1.5	8083	4	US-09-383-630-4
9	36	1.5	8083	4	US-09-383-630-5
10	35.6	1.5	2793	1	US-08-209-747-1
11	35.6	1.5	2793	1	US-08-458-298-1
12	35.2	1.5	2338	1	US-08-425-069-1
13	35.2	1.5	2338	2	US-08-317-844B-1
14	34.8	1.5	1326	4	US-09-457-046B-53
15	34.8	1.5	18596	4	US-09-318-448-11
16	34.6	1.5	289	4	US-09-007-005-17
17	34.6	1.5	289	4	US-09-244-796-17
18	34.4	1.5	216	1	US-07-609-716-45
19	34.4	1.5	216	3	US-08-475-411A-45
20	34.4	1.5	216	4	US-08-478-028A-45
21	33.8	1.4	7218	1	US-08-232-463-14
22	33.6	1.4	1416	3	US-08-911-853-3
23	33.6	1.4	1416	4	US-09-479-409-3
24	33.6	1.4	1416	4	US-09-479-453-3
25	33.6	1.4	2509	2	US-09-014-969-1
26	33.6	1.4	4377	4	US-08-911-853-28
27	33.6	1.4	4377	4	US-09-479-409-28

C 28	33.6	1.4	4377	4	US-09-479-453-28	Sequence 28, Appl
C 29	33.6	1.4	6234	4	US-08-971-188-1	Sequence 1, Appl
C 30	33.4	1.4	1505	1	US-07-915-246-1	Sequence 1, Appl
C 31	33.4	1.4	16442	3	US-08-781-891-208	Sequence 208, App
C 32	33	1.4	2561	4	US-09-270-542-101	Sequence 101, App
C 33	33	1.4	2561	4	US-09-270-542-119	Sequence 119, App
C 34	33	1.4	34303	2	US-08-735-609-4	Sequence 4, Appl
C 35	33	1.4	34303	2	US-08-735-609-4	Sequence 4, Appl
C 36	33	1.4	34303	3	US-09-315-372-4	Sequence 4, Appl
C 37	33	1.4	34303	3	US-09-245-497-4	Sequence 4, Appl
C 38	33	1.4	34303	3	US-09-562-919-4	Sequence 4, Appl
C 39	33	1.4	34303	2	US-08-374-483-6	Sequence 6, Appl
C 40	33	1.4	34308	4	US-08-973-334-3	Sequence 3, Appl
C 41	33	1.4	35408	4	US-09-563-869A-3	Sequence 3, Appl
C 42	33	1.4	35408	4	US-08-549-489-3	Sequence 3, Appl
C 43	33	1.4	35935	2	US-08-735-609-1	Sequence 1, Appl
C 44	33	1.4	35935	2	US-08-735-609-1	Sequence 1, Appl
C 45	33	1.4	35935	2	US-08-735-609-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-846-234-4
Sequence 4, Application US/08846234
Patent No. 6166292
GENERAL INFORMATION:
APPLICANT: OSUMI Chieko
APPLICANT: NOZAKI Jinshi
TITLE OF INVENTION: KIDA Takao
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (Epo)
CURRENT APPLICATION DATA: US/08/846, 234
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON
REGISTRATION NUMBER: 24,618
TELEPHONE: (703)-413-3000
TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2517 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: cucumber (Cucumis sativus)
FEATURE:
NAME/KEY: CDS
LOCATION: 56..2407
US-08-846-234-4
Query Match 100.0% Score 2352; DB 4; Length 2517;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2352; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGGCTCTAGTATTAAAAATGSGGTGCTCCAAAGTAGTTTCATTGTAGTGCTTAAATGAC 60
   |||||
Db 56 ATGGCTCTAGTATTAAAAATGSGGTGCTCCAAAGTAGTTTCATTGTAGTGCTTAAATGAC 115
QY 61 ATGTGCTACCGTTTGCAATCGACGAGTGGATTTCAGTGTGAAGGTCATTCGTTTCTG 120
   |||||
Db 116 ATGTGCTACCGTTTGCAATCGACGAGTGGATTTCAGTGTGAAGGTCATTCGTTTCTG 175
QY 121 TCCGAGTTCCTAGAAACATGTTGCTTCCTCCCTCCGACACTTCGATGAGCAAGTCC 180
   |||||
Db 176 TCCGAGTTCCTAGAAACATGTTGCTTCCTCCCTCCGACACTTCGATGAGCAAGTCC 225
QY 181 CCGGTTTCGGTGGTTCCTTTGTTGATTCGACGCGTCGGAACTGATAGCCGACATGTT 240
   |||||
Db 236 CCGGTTTCGGTGGTTCCTTTGTTGATTCGACGCGTCGGAACTGATAGCCGACATGTT 295
QY 241 GTTTCATTTGGAGAGCTGAAGATATTCGTTATGAGTATTTTCAGGTTTAAAGTTTGG 300
   |||||
Db 296 GTTTCATTTGGAGAGCTGAAGATATTCGTTATGAGTATTTTCAGGTTTAAAGTTTGG 355
QY 301 TGGACTACACACTGGGTGGTCCGAATGGTGGGATCTTGAATCGGAGACTCGATTTGTG 360
   |||||
Db 356 TGGACTACACACTGGGTGGTCCGAATGGTGGGATCTTGAATCGGAGACTCGATTTGTG 415
QY 361 ATCCTTGAGAAGTCAGATTCGTGTCGACCGTATGTTTCCCTCTTCGATCGTTGAAGGA 420
   |||||
Db 416 ATCCTTGAGAAGTCAGATTCGTGTCGACCGTATGTTTCCCTCTTCGATCGTTGAAGGA 475
QY 421 CCGTTCCGAACCTCGATTCAGCTGGGGATGATGACTTTTCGATGTTTGTGTCGAGACT 480
   |||||
Db 476 CCGTTCCGAACCTCGATTCAGCTGGGGATGATGACTTTTCGATGTTTGTGTCGAGACT 535
QY 481 GGTTCCTCGAAAGTTTGTGATGATGATCGTTCGAAAGTATGTTGATCTTCATCGTGGAT 540
   |||||
Db 536 GGTTCCTCGAAAGTTTGTGATGATGATCGTTCGAAAGTATGTTGATCTTCATCGTGGAT 595
QY 541 GATCGTTTGACATGTTTAAAGAGCGGATGAGATCGTGTGAGACCCATCTTGGAACTTTT 600
   |||||
Db 596 GATCGTTTGACATGTTTAAAGAGCGGATGAGATCGTGTGAGACCCATCTTGGAACTTTT 655
QY 601 CGCTTTTGGAGAGAGAACTCCACGAGTATGTTGAGCAAAATTCGGTTGTCACAGTGG 660
   |||||
Db 656 CGCTTTTGGAGAGAGAACTCCACGAGTATGTTGAGCAAAATTCGGTTGTCACAGTGG 715
QY 661 GACGCTTTTACCTAAACGTTTCATCCACAGGGCGTAAATAGAAAGCGCTGAGGCACTCTGTC 720
   |||||
Db 716 GACGCTTTTACCTAAACGTTTCATCCACAGGGCGTAAATAGAAAGCGCTGAGGCACTCTGTC 775
QY 721 GACGGCGTTTCTCCCGTTTACGTCCTATTCGACGATGCTGGCAATCCATCGAGAC 780
   |||||
Db 776 GACGGCGTTTCTCCCGTTTACGTCCTATTCGACGATGCTGGCAATCCATCGAGAC 835
QY 781 GATTGGATCCCTACCAAGAAAGATGAACCAAAACGTCGCGGCGGCAAAATGCCC 840
   |||||
Db 836 GATTGGATCCCTACCAAGAAAGATGAACCAAAACGTCGCGGCGGCAAAATGCCC 895
QY 841 TGGCGTCTTTGAAATTCCAAGAAATTAACAATTCCTGTCAGTACTCAATCCCAAGGCC 900
   |||||
Db 896 TGGCGTCTTTGAAATTCCAAGAAATTAACAATTCCTGTCAGTACTCAATCCCAAGGCC 955
QY 901 ACCGCGCCCCGAGCGCGGCAAGAAAGGATGAGGCGTTTATAGATGAATCAAGAGAGAG 960
   |||||
Db 956 ACCGCGCCCCGAGCGCGGCAAGAAAGGATGAGGCGTTTATAGATGAATCAAGAGAGAG 1015
QY 961 TTTTAAAGCTGTGAGCATGTTTATGTTTGGCATGCTTTGTGTGGATATTTGGGTGGCCTT 1020
   |||||
Db 1016 TTTTAAAGCTGTGAGCATGTTTATGTTTGGCATGCTTTGTGTGGATATTTGGGTGGCCTT 1075
QY 1021 CGCGCCGAGGTGCTGCTGCTGCTGAGGCACTGTGATTCAGCCAGTGGCTTTACACAGAG 1080
   |||||
Db 1076 CGCGCCGAGGTGCTGCTGCTGCTGAGGCACTGTGATTCAGCCAGTGGCTTTACACAGAG 1135
QY 1081 CTGACAGATGACATGAGAGATTGTCGCGTGAGATAGATTTGTTCTTCATAAGTTCGGGCTG 1140
   |||||

Db 1136 CTGACAGATGACATGAGAGATTGTCGCGTGATAGATTTGTTCTTCATAAGTTCGGGCTG 1195
QY 1141 GTCCCGCCGAGAGAGGCTGAGGAGATGTACGAAGGACTTCATCTGATTTGGAAAAAGTT 1200
   |||||
Db 1196 GTCCCGCCGAGAGAGGCTGAGGAGATGTACGAAGGACTTCATCTGATTTGGAAAAAGTT 1255
QY 1201 GGGATGACGCGTGTAAAGATTGACGTTATCCACTATTTGAGATGTTGTTGTAAGACTAT 1260
   |||||
Db 1256 GGGATGACGCGTGTAAAGATTGACGTTATCCACTATTTGAGATGTTGTTGTAAGACTAT 1315
QY 1261 GGAAGGAGAGTGAATTTGGCAAGGCAATATTACAAAGCAATGACCAATCAATTAATAA 1320
   |||||
Db 1316 GGAAGGAGAGTGAATTTGGCAAGGCAATATTACAAAGCAATGACCAATCAATTAATAA 1375
QY 1321 CATTTTAAAGAAATGAGATTCATGCAAGATGAGACATGTATAGCACTCATGTTCTCT 1380
   |||||
Db 1376 CATTTTAAAGAAATGAGATTCATGCAAGATGAGACATGTATGCAAGCATTCATGTTCTCT 1435
QY 1381 GGCACGAAAGCTATCTCTCTGTCGTGTCGTGATGACACTTTTGGTGCAGGACCCCTCT 1440
   |||||
Db 1436 GGCACGAAAGCTATCTCTCTGTCGTGTCGTGATGACACTTTTGGTGCAGGACCCCTCT 1495
QY 1441 GGTGATCCAAAGCGTACGTTTGGCTCCAAAGATGTACATGTTGATTTGTCGCAAGAC 1500
   |||||
Db 1496 GGTGATCCAAAGCGTACGTTTGGCTCCAAAGATGTACATGTTGATTTGTCGCAAGAC 1555
QY 1501 AGCTTTGATGAGTGGGAACTTCATCCACCCTGACGAGATGTTGTTCAATCCACCCACT 1560
   |||||
Db 1556 AGCTTTGATGAGTGGGAACTTCATCCACCCTGACGAGATGTTGTTCAATCCACCCACT 1615
QY 1561 TGTGCCGCTTCATGCTGCTGCTTCGAGGCAATCTGTTGGGCCGATCTATGTTAGTAT 1620
   |||||
Db 1616 TGTGCCGCTTCATGCTGCTGCTTCGAGGCAATCTGTTGGGCCGATCTATGTTAGTAT 1675
QY 1621 TCTGTGGGAAAGCATTAATTTGATCTTCTGTAATAAACTAGTCTGCTGATGATGATC 1680
   |||||
Db 1676 TCTGTGGGAAAGCATTAATTTGATCTTCTGTAATAAACTAGTCTGCTGATGATGATC 1735
QY 1681 CTTCGAAAGTATGATGATGATCCGAGCTCGGATGTTGTTGTAAGACCTTTTGAT 1740
   |||||
Db 1736 CTTCGAAAGTATGATGATGATCCGAGCTCGGATGTTGTTGTAAGACCTTTTGAT 1795
QY 1741 AATGGAAGAACTATGCTTAAATTTGAATTTGAATCTGACAAAGTTCACTGAGTATGTGCA 1800
   |||||
Db 1796 AATGGAAGAACTATGCTTAAATTTGAATTTGAATCTGACAAAGTTCACTGAGTATGTGCA 1855
QY 1801 TTTCATCTGCCAAGAGAGAGATGTTGTCTGTGAAGACAGCCGCAACCAATGCTTTGACAA 1860
   |||||
Db 1856 TTTCATCTGCCAAGAGAGAGATGTTGTCTGTGAAGACAGCCGCAACCAATGCTTTGACAA 1915
QY 1861 TACTCAAAAAGATGACATCAAAACCTAAACCCAAAGACATAGAAATGAGCAGATGAGAA 1920
   |||||
Db 1916 TACTCAAAAAGATGACATCAAAACCTAAACCCAAAGACATAGAAATGAGCAGATGAGAA 1975
QY 1921 AACCTATCTCTATTTGAAGGCGTTTAAACCTTTCGCTTTTACCTGTATCAAGCCAAAAA 1980
   |||||
Db 1976 AACCTATCTCTATTTGAAGGCGTTTAAACCTTTCGCTTTTACCTGTATCAAGCCAAAAA 2035
QY 1981 CTTATCTCTCCAAAGCCCTCTCAAGATCTTGAACATAGCTTTGAACCATTCGAAATTCGAG 2040
   |||||
Db 2036 CTTATCTCTCCAAAGCCCTCTCAAGATCTTGAACATAGCTTTGAACCATTCGAAATTCGAG 2095
QY 2041 CTTCATCACTGTTTACACAGTGAACCAATCTCAACAACTCTCTACACTTTGCCCCAAAT 2100
   |||||
Db 2096 CTTCATCACTGTTTACACAGTGAACCAATCTCAACAACTCTCTACACTTTGCCCCAAAT 2155
QY 2101 GGGCTGTGTAACATGCTTAAACACTAGTGAAGCCATCCAAATCTGTGATGATGACATGAC 2160
   |||||
Db 2156 GGGCTGTGTAACATGCTTAAACACTAGTGAAGCCATCCAAATCTGTGATGATGACATGAC 2215
QY 2161 CTTAAAGTCAAGTGAAGTTGGTGTCAAAAGGCTGTGTGAGATGCGAGTATTTGATCGAAA 2220
   |||||
```


Db 2216 CTAAGCTCAGTCGAGATGTCGTCAAAAGGCTGTGTCGATGCCAGTATTTTCATCGAANA 2275
Oy 2221 AAACCAAGGCTTGTCTGATTTGATGGAGAGATGTTGGTTCAAGTATGATCAGACCAA 2280
Db 2276 AAACCAAGGCTTGTCTGATTTGATGGAGAGATGTTGGTTCAAGTATGATCAGACCAA 2335
Oy 2281 ATGTGTGTGTTCAAGTGGCATGGCCAAATTGATTCCTATCGGGTGGCATTTTCGGTTATC 2340
Db 2336 ATGTGTGTGTTCAAGTGGCATGGCCAAATTGATTCCTATCGGGTGGCATTTTCGGTTATC 2395
Oy 2341 GAGTACTTCTTT 2352
Db 2396 GAGTACTTCTTT 2407

RESULT 2

US-09-561-825-23
; Sequence 23, Application US/09561825
; Patent No. 6337450
; GENERAL INFORMATION:
; APPLICANT: ISHIGE, Fumiharu
; APPLICANT: MATSUNABE, Eijiro
; APPLICANT: OEDA, Kenji
; TITLE OF INVENTION: PLANT PROMOTERS
; FILE REFERENCE: 2185-0431P
; CURRENT APPLICATION NUMBER: US/09/561.825
; CURRENT FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: JP 11-124527
; PRIOR FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: JP 11-247211
; PRIOR FILING DATE: 1999-09-01
; NUMBER OF SEQ ID NOS: 30
; SEQ ID NO 23
; LENGTH: 2498
; TYPE: DNA
; ORGANISM: Glycine max cv. Williams 82
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (62)..(2407)
US-09-561-825-23

Query Match 41.5%; Score 976.4; DB 4; Length 2498;
Best Local Similarity 66.0%; Pred. No. 3,3e-305;
Matches 1507; Conservative 0; Mismatches 741; Indels 36; Gaps 5;

Oy 57 TCACATGTCGACCCCTTTCGAATCGACGATTCGATTCCTGCTGTAACGGTCATTCGTT 116
Db 118 TAATTTGCTTTGCTCCATTAACCTTGAAGATCAATTTCTCTGCGCAAGGCCACCTTT 177
Oy 117 TCTGTCCGATGTTCTGAGAACATTTGCTTCTCTCTCCGTAC-----AC 164
Db 178 TCTCAGGAAAGTTCCGAAACATATATCTACACCTTACCCATGACGCGCAAGAGTAG 237
Oy 165 TTCGATAGACAAGTCCCGGTTTGGTGGTCTTGTGTTGATTCGACGCGTCGGAACC 224
Db 238 TAAGAACACGAGGAGACGAGCTGTAGTTCCTCTGCGCTTCACGCGGACGAGCC 297
Oy 225 TGATACCGACATGTTGTTTCATTTGGGAAGCTGAAGATATTGGTTATATGATTTT 284
Db 298 CAGAAAGCCGACGTCGTTCCCTGGGAAAGCTCAGAGAAATAAATTCATGACATATT 357
Oy 285 CAGTTTAAAGTTTGGTGTGACTACACACTGGTGGTGTGCAAAATGTTGGGATTTGAATC 344
Db 358 CCGGTTTAAAGTGTGTGTGACACACTGCTGCGGTGTAGCAACGACGAACTGAGCA 417
Oy 345 GGAGACTCAGATTTGATCCTTGAGAGTCAAG---ATTCTGTGACCGGTATGTTTCT 401
Db 418 CGAGACACAGATGATCTCTCGACAAAAACGACGACTCGGACGCCCTTGTGTGAT 477
Oy 402 TCTTCCGATCGTTGAGGACGCTTCCGAACCTCGATTCAGCCGTTGGGATGATCTTGT 461
Db 478 TCTCCGATCTCCAAAGCTCTGTTCCGAGCTCTCCGCAACCCGTTTGGATGATTAAGC 537

Oy 462 CGATGTTTGTCTGAGAGTGTGTCGAAAGTGTGATGATGATCGGAATGATTT 521
Db 538 GGACGTTTTCATGAGAGCGGGTGCACAGCTGTCTGTGCTCCACTTGGGAGCTGCTT 597
Oy 522 GTATCTCATGCTGTGATGATCCGTTTCCACTTGTTAAGAGCGGATTAAGATCTGAG 581
Db 598 ATACGTCACGTTTGGCGATGAGCCCGTATCAGTTGCTTAAGAAAGCAATTAAGTCTGTAG 657
Oy 582 GACCCATCTTGGAACTTTTCCGTTGTTGGAGGAAGAAGATCTCACACAGTATGCTGACAA 641
Db 658 GATGCAATTTGGGACCTTAACCTTCTCGAGGAGAAACCGCGCATGATCATAGACAA 717
Oy 642 ATTGGTTGTGTCACGTGGAGCGCTTTTACCTAACGTTTCAACAGGCGCTAATAG 701
Db 718 GTTGTGTTGTGATACATGAGGACCGCTTTTACTTGAAGTGTGATCCCTCAGGTGTGGGA 777
Oy 702 AGCGGTGAGCATCTCTGTGACGCGGCTGTCTCCCGTTTAACTTAACTGATGATGATG 761
Db 778 AGGGGTGAAGGGTGTGTGGAGGAGGAGGTCCTCCAGGATGTCTTAATGACAGCGG 837
Oy 762 TTGGCAATCCATCGACAGATTCGATTCCTATCAC---CAAGAAGAAATGAACCAAC 818
Db 838 GTGGCAAGCATTTTGTACAGAGAGAGAGAGCCCATACGAGCAAGAGGCTATGAGCGAAC 897
Oy 819 CGTGGCGGCGAGCAAAATGCCCTGCTCTTTTGAATTCAGAGAAATTAACAATTCG 878
Db 898 CTCGCCAGGCGCAAAATGCCATGAGTTGTGTAAGTTGAGAGAAATTAACAAGTTTCA 957
Oy 879 TGACTATGTCATCCCAAGGCCACCGCCCGAGCGGCGGCAAGAGGATGAAGCGCTT 938
Db 958 ACAGTATTTAGTGAAGAGATTTCTG-----AGAAAGGATATGGTCTCTT 1002
Oy 939 TATAGATGAACCTCAAGAGAGATTTAAGACTGTGAGCATGTTTATGTTTGCATCTTT 998
Db 1003 TGTTAGGACATTTGAAGAACAGATTTAGAGCGTGGAGAGATGTATGTGTGACCGCT 1062
Oy 999 GTGTGATATTTGGGCTGAGCTTTCGCGGAGAGTGCCTGCTTGCCTGAGGACGTTGAT 1058
Db 1063 TGTGTGGATTTGGGCTGAGGCTTTCAGACCAAGGTTCCGGGATCCCAAGGCTTAAGTGT 1122
Oy 1059 TCAGCCAGTGTCTTTCACAGGCTCGACATGACGATGAGAGATTTGGGCTGATTAAGT 1118
Db 1123 CACTCGGAAGCTGTCAAAATGAGACTAAATTTGACAAAGAAAGATTTAGCGTGAATTAAGT 1182
Oy 1119 TGTCTTCAATPAGGTCGGGCTGTCCCGGAGAGGCTGAGAGATGTAGCAAGAGCT 1178
Db 1183 GGTCACTTAACGAGTGTGACTGTGACACACCTGTGCTCACCTTTTGTAGAGGGGCT 1242
Oy 1179 TCATGCTCATTTTGAAGAAAGTTGGATTCGACGCTGTTAAGATTTGACCTTATTCACCTATT 1238
Db 1243 CCACCTCCGTTTGGATTCGTGGGCTATTGACGCTGTTAAGTGTGACCTTATTCACCTTGT 1302
Oy 1239 GGAGATGTTGTGGAAGCATGTGAGGAGAGTGTGATTTGGCAAGGCATATTACAAGC 1298
Db 1303 CGAGATGCTATTCGAGAGATACGCTGCGGTGTGAGCTAGCCAAAGCTTATTACCAAGC 1362
Oy 1299 AATGACCAATCAATTAATAAATTAATTAAGGAATGAGATTCATTCGAAGTATGAACA 1358
Db 1363 GCTCACTGCTTGGTCAAGAGCATTTCAAGGCAATGGGCTATTGGCAGCATGAGCA 1422
Oy 1359 TTGTAAAGATTCATGTTCTTGGCAAGAGCTTATCTCTTGTGCTGTTGTGATGA 1418
Db 1423 TTGTAAATGACTTCTTCTCTGTTACGCAAGCATTAAGCCCTTGGGCGGTAGAGATGA 1482
Oy 1419 CTTTGTGTGACGAGACCCCTCTGTGTATCAAAAGTGTGATTTGGCTCCAAGATGTCA 1478
Db 1483 TTTTGTGTGACAGATCCCTCTGAGATCAAAATGACAGCTATTGGCTCAAGAGGTGTA 1542
Oy 1479 CATGTTTCAATGAGCAAGACGCTTGTGAGTGGGAACTTCACTACACCTACTGTGGA 1538
Db 1543 CATGTGTGACGTGTGCTTACACAGCTTGTGATGGGAATTTTATTCAGCGGATTTGGGA 1602
Oy 1539 TATGTTCAATCCACCAACCCCTTGTGCGGCTTTCATGTGCTCTTGAAGCATCTCTGG 1598

```

Db 1603 CATGTTCCAGTCACCTGCTGTCGCAATTCATGACGCTTAGGGCCATCTGTG 1662
Qy 1599 TGGCCGATCTATGATGATCTGTTGGGAAGCATATCTGATCTCTGAAAAACT 1658
Db 1663 TGGACCATGTTACGTTGATGATGTTGGGAAGCATATCTGATCTCTGAAAAACT 1722
Qy 1659 AGTGTCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1718
Db 1723 CCTTGTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1782
Qy 1719 TTTGTTTGAAGACCTTTCATATGAGAACTATGCTTAAGATTGGAATCTCAACA 1778
Db 1783 TTTGTTTGAAGACCTTTCATATGAGAACTATGCTTAAGATTGGAATCTCAACA 1842
Qy 1779 GTTCACTGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1838
Db 1843 ATATACAGGTGTTTGGGCTGCTATTTATTTATTTGCAAGAGAGTGGTGTGCTG 1902
Qy 1839 CCGCAACCAATCTTTTCAATACTCAATACTCAATACTCAATACTCAATACTCA 1898
Db 1903 GAGAAACAAGAGTGGCTGCTGATTTTCACTGATGATGATGATGATGATGATG 1962
Qy 1899 CATGATATGAGCAGTGGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1958
Db 1963 CATGATATGAGCAGTGGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2022
Qy 1959 TTACCTCTATCAAGCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2018
Db 2023 ATATTTGTTCAAGAGCAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2082
Qy 2019 TCTTGACCATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2078
Db 2083 ACTTATGAGCATTTACTTTGATTTGATTTGATTTGATTTGATTTGATTTGAT 2142
Qy 2079 TTTCTACACTTTGCCCCAATTTGGGCTGCTGATTTGATTTGATTTGATTTGAT 2138
Db 2143 GTTATTTCAATTTGCTCCATTTGATTTGATTTGATTTGATTTGATTTGATTT 2202
Qy 2139 ATCTGTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2198
Db 2203 GTTCATGAGTTTGTGACAAACCATATGATGATGATGATGATGATGATGATG 2262
Qy 2199 GATGCGATATTTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2258
Db 2263 GATGAGGATTTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2322
Qy 2259 GTTCAAGTATGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 2318
Db 2323 ATTGATTTATG---AGGATTTATGCTGAGAGTGCAGATTCCCTGCTGATG 2379
Qy 2319 ATCG 2322
Db 2380 ATTG 2383

```

RESULT 3
 US-08-232-463-14/c
 : Sequence 14, Application US/08232463
 : Patent No. 5670367

: GENERAL INFORMATION:
 : APPLICANT: DORNER, F.
 : APPLICANT: SCHEIFLINGER, F.
 : APPLICANT: FALKNER, F. G.
 : TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
 : NUMBER OF SEQUENCES: 52
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: Foley & Lardner
 : STREET: 1800 Diagonal Road, Suite 500
 : CITY: Alexandria
 : STATE: VA
 : COUNTRY: USA
 : ZIP: 22313-0299

```

: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232,463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935,313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)856-9300
: TELEFAX: (703)683-4109
: TELEX: 899149
: INFORMATION FOR SEQ. ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgpt-Fls
US-08-232-463-14

```

Query Match 2.0%; Score 47.4; DB 1; Length 7218;
 Best Local Similarity 4.7%; Pred. No. 0.00024;
 Matches 15; Conservative 178; Mismatches 124; Indels 0; Gaps 0;

```

Qy 1077 AGGCGTCGATGAGCATGAGGATTTGGCGGTGATAGATTGTTCTTCATAGAGTGG 1136
Db 1356 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1297
Qy 1137 GCTGTCGCCGAGAGAGAGTGGAGAGATGACGATTCATTCATTTGGAAAA 1196
Db 1296 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1237
Qy 1197 AGTTGGATGACGCGTTTAACTTGCCTTATCCACTATTGAGATTTGTGAGAGA 1256
Db 1236 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1177
Qy 1257 CTAATGAGGAGAGTGGATTTGGCAAGGCATATTACAAGCATGACCAATCAATA 1316
Db 1176 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1117
Qy 1317 TAAACATTTTAAAGAAATGAGATTCATTCAGATATGAGAACTTGAACACTTCAT 1376
Db 1116 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1057
Qy 1377 CCTTGGCAGGAACTTA 1393
Db 1056 CCTGACCTGCAGCCA 1040

```

RESULT 4
 US-09-457-046B-19
 : Sequence 19, Application US/09457046B
 : Patent No. 6287835
 : GENERAL INFORMATION:
 : APPLICANT: Croteau, Rodney et al.
 : TITLE OF INVENTION: Transcylases of the Facilitaxel Biosynthetic Pathway
 : FILE REFERENCE: 53679
 : CURRENT APPLICATION NUMBER: US/09/457,046B
 : CURRENT FILING DATE: 1999-12-07
 : NUMBER OF SEQ ID NOS: 74
 : SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 19
LENGTH: 911
TYPE: DNA
ORGANISM: Taxus cuspidata
US-09-457-046B-19

Query Match 1.6%; Score 38; DB 4; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.065;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

OY 317 TTGGTCGGAATGGTGGGATCTTGGAATCGGAGACTCAGATTGTGCTTCTTGGAAGTCAAG 376
DB 461 TTGGTATTAAGTTGACACCATAGATATATGAACAATGCTTATGAGAGAAATGTAATG 520
OY 377 ATTCTGTCGACCGATGATTTCTTCTTCCGATGCTTGAGGAGCGTTCCGAACTCGA 436
DB 521 ATTTTGTTCGTCCTTTCAGTAGCGAGCATTTGCTTGGATGACCAAGAGCAAGGCTC 580
OY 437 TTCACCGCTGGGAGATGATCTTTCGATGTTTGTGCGAGAGTGTTCGTCGAAGTTG 496
DB 581 TTCAATTCACATACATGAGATGGAAGCTTCCTTTCGATGATTTGAGAGAAATATAT 640
OY 497 TTGATGATCGTTCGGAAGTATGTTGATCTTCATGCTGCTGATGATCCGTTGCACCTG 556
DB 641 TTAATCCCCCACTTCCAATGATATATGTAATGCCATTGGTACTGCATATGCAATGG 700
OY 557 TTAAG 562
DB 701 ATAAATG 706

US-09-457-046B-21

Query Match 1.6%; Score 38; DB 4; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.065;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

OY 317 TTGGTCGGAATGGTGGGATCTTGGAATCGGAGACTCAGATTGTGCTTCTTGGAAGTCAAG 376
DB 461 TTGGTATTAAGTTGACACCATAGATATATGAACAATGCTTATGAGAGAAATGTAATG 520
OY 377 ATTCTGTCGACCGATGATTTCTTCTTCCGATGCTTGAGGAGCGTTCCGAACTCGA 436
DB 521 ATTTTGTTCGTCCTTTCAGTAGCGAGCATTTGCTTGGATGACCAAGAGCAAGGCTC 580
OY 437 TTCACCGCTGGGAGATGATCTTTCGATGTTTGTGCGAGAGTGTTCGTCGAAGTTG 496
DB 581 TTCAATTCACATACATGAGATGGAAGCTTCCTTTCGATGATTTGAGAGAAATATAT 640
OY 497 TTGATGATCGTTCGGAAGTATGTTGATCTTCATGCTGCTGATGATCCGTTGCACCTG 556
DB 641 TTAATCCCCCACTTCCAATGATATATGTAATGCCATTGGTACTGCATATGCAATGG 700
OY 557 TTAAG 562
DB 701 ATAAATG 706

RESULT 6
US-09-740-035-3/C
Sequence 3, Application US/09740035
Patent No. 6344353
GENERAL INFORMATION:
APPLICANT: YE, Jane et al
TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
FILE REFERENCE: C1001058
CURRENT APPLICATION NUMBER: US/09/740,035
CURRENT FILING DATE: 2000-12-20
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 19736
TYPE: DNA
ORGANISM: Human
US-09-740-035-3

Query Match 1.6%; Score 36.8; DB 4; Length 19736;
Best Local Similarity 53.5%; Pred. No. 1.3;
Matches 77; Conservative 0; Mismatches 67; Indels 0; Gaps 0;

OY 1809 CCAAGGAGAGATGTCGTCGTCGACACAGCCGCAACCAATGCTTTCACATACCTCAA 1868
DB 2368 CCTGGAGAGTGGAGTGTGTAGTACCTAAGACTGCACCTGACCTCGGCGCAGC 2309
OY 1869 ACAGTGCATCCAACTAACTAACCCAAAGACATAGATGCGACACTGGAGAAACCTAT 1928
DB 2308 AGAGGAAGACTCCTGCTCAAAAACAAAGAAAGAAAGAGAGGTGAGTCCCTGT 2249
OY 1929 CTCTATTGAGGCGTTAAACCTT 1952
DB 2248 CCTCAGTGAAGAGCTTACAGCCT 2225

US-09-010-928B-3

Sequence 3, Application US/09010928B
Patent No. 5994099
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V
APPLICANT: Hayashi, Cheryl Y
TITLE OF INVENTION: EXTREMELY ELASTIC SPIDER SILK PROTEIN AND DNA
TITLE OF INVENTION: CODING THEREFOR
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESSES:
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH
STREET: 8110 GATEHOUSE RD. SUITE 500E
CITY: FALLS CHURCH
STATE: VIRGINIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 22042
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/010,928B
FILING DATE: 22-JAN-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M
REGISTRATION NUMBER: 28977
REFERENCE/DOCKET NUMBER: 1447-109P
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2824 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant

[illegible]

```

RESULT 10
US-08-209-747-1
: Sequence 1, Application US/08209747
: Patent No. 5733771
: GENERAL INFORMATION:
: APPLICANT: Lewis, Randolph V.
: APPLICANT: COLGIN, Mark
: TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
: TITLE OF INVENTION: Silk Proteins
: NUMBER OF SEQUENCES: 56
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Birch, Stewart, Kolasch & Birch
: STREET: P.O. Box 747
: CITY: Falls Church
: STATE: Virginia
: COUNTRY: USA
: ZIP: 22040-3487
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/209,747
: FILING DATE: 14-MAR-1994
: CLASSIFICATION: 530
: ATTORNEY/AGENT INFORMATION:
: NAME: Murphy Jr., Gerald M.
: REGISTRATION NUMBER: 28,977
: REFERENCE/DOCKET NUMBER: 1447-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2793 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHEICAL: NO
: ORIGINAL SOURCE:
: ORGANISM: Nephila clavipes
: TISSUE TYPE: minor ampullate gland
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 183..2675
: OTHER INFORMATION: /product= "N. clavipes minor
: OTHER INFORMATION: ampullate silk protein"
US-08-209-747-1

Query Match 1.5%; Score 35.6; DB 1; Length 2793;
Best Local Similarity 52.7%; Pred. No. 0.84;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

```

[illegible]

```

US-08-458-298-1
US-08-458-298-1
Sequence 1, Application US/08458298
Patent No. 5756677
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Colgin, Mark
TITLE OF INVENTION: cDNAs Encoding Minor Ampullate Spider
NUMBER OF SEQUENCES: 56
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolach & Birch
STREET: P.O. Box 747
CITY: Falls Church
STATE: Virginia
COUNTRY: USA
ZIP: 22040-3487
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,298
FILING DATE: 02-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/209,747
FILING DATE: 14-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Murphy Jr., Gerald M.
REGISTRATION NUMBER: 28,977
REFERENCE/DOCKET NUMBER: 1447-104P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-205-8000
TELEFAX: 703-205-8050
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2793 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Nephila clavipes
TISSUE TYPE: minor ampullate gland
FEATURE:
NAME/KEY: CDS
LOCATION: 183..2675
OTHER INFORMATION: /product= "N. clavipes minor
OTHER INFORMATION: ampullate silk protein"
US-08-458-298-1

```

Query Match	1.5%	Score 35.6	DB 1	Length 2793
Best Local Similarity	52.7%	Prod. No. 0.84		
Matches	77	Conservative	0	Mismatches 69; Indels 0; Gaps 0;
QY	1029	GGTGGCTGGCTGCTGCAGGACATGTGTATGACGAGTCTTCCAGCGGGCTGCAGAT	1088	
DB	1679	GCTGCCGAGCAGACAGCTGTGCTGGCTCTGCTCTGTCGACAGACAGACAGATCTGTGT	1738	
QY	1089	GACGATGGAGAGATTTGGCCGGCTGGATTAAGATTGTTCTTCAATTAAGTTCGGGCTGTCGCCGC	1148	

Db 1739 TACGGTAGAGGTGCTGGTGGTGGAGCTGGAGAGCTGCTGGAGGTGCTGGAGCTGGAGCCGCT 1798

QY 1149 GGAGAGGCTGAGGAGATGTACGAAG 1174

Db 1799 GCAGGTGCAGAGACGAGGAGCTGGAGG 1824

RESULT 12
HS-08-A25

Sequence 1, Application US/08425069
Patent No. 5728810
GENERAL INFORMATION:
APPLICANT: Lewis, Randolph V.
APPLICANT: Xu, Ming
APPLICANT: Himan, Michael B.
TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Birch, Stewart, Kolasch & Birch
STREET: 301 No. 5728810th Washington Street
CITY: Falls Church
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22046

Query Match	1.58;	Score 35.2;	DB 1;	Length 2338;
Best Local Similarity	49.08;	Pred. No. 1;		
Matches 94; Conservative	0;	Mismatches 98;	Indels 0;	Gaps 0

QY	1028	AGGTGCTGGCTTGGCCGAGGCAACGTGTAATTCACGCAATGCTTTACACAGAGCCGACGA	1087
Db	1281	AGGTGGATTAGGTGGACAAAGGGCAGGTGCACAGCCGTGCAGCAGCTGGAGAGTGGCCG	1340
QY	1088	TGACGATGGAGGATTTGGCCGGTGATTAAGATTCCTTCATTAAGGTGGGGCTGGTCCCGC	1147
Db	1341	ACAAGGAGGATTAGGAGGCCCTTGAAACCAGAGCTGTGACGAGGTGGACAAAGGTGCAGC	1400
QY	1148	CGGAGAAAGCTGAGGAGATTTACGAAGGACCTTCATGCTCATTTTGGAAAAAGTTGGGATCG	1207
Db	1401	AGCAGCAGCTGGAGAGGTCCGCGACAAAGAGATATGGAGGTCTTTGGAAACCAAGGTGCTGGC	1460
QY	1208	ACGGGTTTAAGA	1219
Db	1461	ACGAGGTGACAA	1472

RESULT 1.3
116-00-317

/ Sequence 1, Application US/08317844B
 / Patent No. 5983894
 / GENERAL INFORMATION:
 / APPLICANT: Lewis, Randolph V.
 / APPLICANT: Xu, Ming
 / APPLICANT: Hinman, Michael B.
 / TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
 / TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
 / TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
 / NUMBER OF SEQUENCES: 62
 / CORRESPONDENCE ADDRESS:
 / ADDRESSEE: Birch, Stewart, Kolasch & Birch
 / STREET: 301 No. 5989894th Washington Street
 / CITY: Falls Church
 / STATE: Virginia
 / COUNTRY: U.S.A.
 / ZIP: 22046
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Floppy disk
 / COMPUTER: IBM PC compatible
 / OPERATING SYSTEM: PC-DOS/MS-DOS
 / SOFTWARE: PatentIn Release #1.0, Version #1.25
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/08/317,844B
 / FILING DATE: 04-OCT-1994
 / CLASSIFICATION: 435
 / ATTORNEY/AGENT INFORMATION:
 / NAME: Murphy Jr., Gerald M
 / REGISTRATION NUMBER: 28,977
 / REFERENCE/DOCKET NUMBER: 1447-105P
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (703) 241-1300
 / TELEFAX: (703) 241-2848
 / TELEX: 248345
 / INFORMATION FOR SEO ID NO: 1:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 2338 base pairs
 / TYPE: nucleic acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: cDNA
 / HYPOTHETICAL: NO
 / ORIGINAL SOURCE:
 / ORGANISM: Nephilia clavipes
 / FEATURE:
 / NAME/KEY: CDS
 / LOCATION: 1..2154
 / OTHER INFORMATION: /product= "Nephilia clavipes
 / OTHER INFORMATION: dragline silk protein"
 / PUBLICATION INFORMATION:
 / AUTHORS: Xu, Ming
 / TITLE: Lewis, Randolph V.
 / TITLE: Structure of a protein superfiber: Spider

? TITLE: draftline SLIK
 ? JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
 ? VOLUME: 87
 ? PAGES: 7120-7124
 ? DATE: Sept.-1990
 ? RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 2358
 ? OS-08-317-844B-1

Query Match	1.58;	Score 35.2;	DB 2;	length 2338;
Best Local Similarity	49.08;	Pred. NO. 1;		
Matches 94; Conservative	0;	Mismatches 98;	Indels 0;	Gaps 0;

Oy 1028 AGG1GCCCTGGCTTGGCTTGAGGGCACCTGTGATTACAGCCAGTGCCTTTTCACAGGGGCTGCAGA 1087
 ||||| | ||| | ||| ||||| ||| | ||| |
 Db 1281 AGCTGATTAGTGTGACAAAGGGGCAAGGTGCAGACAGCCGCTGCAGACACTGAGAGGTGCCGG 1340

Oy 1088 TACCGATGAGGATTTGGCGGTGGAATAGATGTTCTTCATTAAGGTGGGCTGGTCCCGC 1147
 | | | | | | | | | | | | | | | | | | | | | |
 Db 1341 ACAAGGAGGATTATGAGGCCCTTGGAAACCAAGTCTCGAGAGAGGTGGACAAAGTGCAGC 1400

0y 1148 CGGAGMGGCTGAGGAGATGTACGAAGACTTCATCTCATTTGGAAAAGTTGGATCG 1207
| | | | | | | | | | | | | | |
Db 1401 AGCAGCACTTGAGAGTCCGACAAGAGAGATATGAGAGTCTTGGAAGCCAAGGTGCTGG 1460

Qy	1208	ACGGTGT	AAGA	1219
Db	1461	ACGAGGTG	GACA	1472

RESULT 14
US-09-457-046B-53
; Sequence 53, Application US/09457046B

INVENTOR: Chishti, Naveed et al.
 TITLE OF INVENTION: Transsialosylases of the Pacilactaxel Biosynthetic Pathway
 FILE REFERENCE: 53679
 CURRENT APPLICATION NUMBER: US/09/457,046B
 CURRENT FILING DATE: 1999-12-07
 NUMBER OF SEQ. ID NOS.: 74
 SOFTWARE: PatentIn Ver. 2.1

```
; LENGTH: 1326
; TYPE: DNA
; ORGANISM: Taxus cuspidata
US-09-457-046B-53
```

Query Match	1.58;	Score 34.8;	DB 4;	Length 1326;
Best Local Similarity	46.38;	Pred. No. 0.92;		
Matches 114; Conservative	0;	Mismatches 132;	Indels 0;	Gaps 0

Oy 317 TTGGTCGAATGGTGGGANTCTTGATCGGAGACTCAGATTGTGATCCTTGAGAAGTCAG 376
||| ||| | ||| ||| ||| ||| |||
Db 686 TTCTTATAACGTTGACACCATTAGAAATATATCAAGCAATGTGTCACTGGAGGAATCTAATG 745

Oy 377 ATTTCGGTCCACCAGATGTTTCTCCTTCCGATCGTTGAGGACCGTTCCGAACCTCGA 436
| | | | | | | | | | | | | |
Db 746 AATTTCGTGCTCTTTTGAACTAGTGCACGATGTTGGATTAGCACGACCAAGSCTC 805

Oy 437 TTCAACCTGGGATGATGACATTGTCATGTTTGTGTGCAGAGTGTTCGTCGAAATTTG 496
 ||| | ||| ||| | ||| | ||| | ||| |
 Db 806 TTCAATTCCACATCTGAGAAATGTGAAGCTTCTCTTTCGATGATTTGAGGAATTTAT 865

Oy 497 TTGATGCATCGTTCCGAAGTATTGTATCTTCATGCTGGTGATCATCCGTTGGACTTG 556
|| || | | | | | | | | | | | | | |
Db 866 TTTAATCCCCCACTTCCAATGGATATTATGTGTAATGCCATTGGTACTGCATATGCATATGG 925

QY	557	TTAAAG	562
Db	926	ATAATG	931

RESULT 15

```

US-09-318-448-11
: Sequence 11, Application US/09318448
: Patent No. 6210950
: GENERAL INFORMATION:
: APPLICANT: Johnson, William G.
: TITLE OF INVENTION: METHODS FOR DIAGNOSING, PREVENTING, AND TREATING
: TITLE OF INVENTION: DEVELOPMENTAL DISORDERS
: FILE REFERENCE: 601-1-057
: CURRENT APPLICATION NUMBER: US/09/318,448
: CURRENT FILING DATE: 1999-05-25
: NUMBER OF SEQ ID NOS: 46
: SOFTWARE: Patentln Ver. 2.0
: SEQ ID NO 11
: LENGTH: 18596
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-318-448-11

```

Query Match	1.5%	Score 34.8;	DB 4;	Length 18596;
Best Local Similarity	57.3%;	Pred. NO. 5.6;		
Matches 63;	Conservative 0;	Mismatches 47;	Indels 0;	Gaps 0;

Dy 1809 CCAAGGAGGAGGATGGTGTCTCGTGACACACGCCGGAACCAATGCTTTTCACAATACTCAA 1868
|| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 5157 CCGGGAGGTGGAAGGCTGCCGTGAGCCACAGCATTCGGCCATTGCACATCCAGCCTGGGCAAC 5216

Qy 1869 ACGAGTGACATCCAAAACATAACCCAAAGA CATGGAATGGCAGAGTGGAG 1918
| | | | | | | | | | | |
Db 5217 AAGAGCGAAACTCTCAAAAAACA AAAAAGGATGGCTTC CATTATGGGTG 5266

Search completed: April 3, 2003, 19:26:12
Job time : 249 secs

```
Search completed: April  3, 2003, 19:26:12
Job time : 249 secs
```


GenCore version 5.1.4.P5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 3, 2003, 18:19:36 ; Search time 248 Seconds
(without alignments)
8318.946 Million cell updates/sec

Title: US-09-675-208-4_COPY_56_2407

Perfect score: 2352

Sequence: 1 ATGCCTCCTACTTTTAAAAA.....CGTTATCGACTTCTTT 2352

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 593429 seqs, 438583890 residues

Total number of hits satisfying chosen parameters: 1186858

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/2/pubpna/PC1_NEW_PUB.seq:*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/2/pubpna/PC10S_PUBCOMB.seq:*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	113.2	4.8	1036	10	US-09-770-445-180
2	86.2	3.7	1056	10	US-09-878-574-2803
3	75.4	3.2	296	10	US-09-294-0938-2999
4	45.8	1.9	377	10	US-09-878-574-1840
5	44.4	1.9	160	10	US-09-770-606-410
6	38	1.6	911	9	US-09-866-570A-19
7	38	1.6	911	9	US-09-866-570A-21
8	38	1.6	911	10	US-09-866-572A-19
9	38	1.6	911	10	US-09-866-572A-21
10	37.6	1.6	46050	10	US-09-820-003A-3
11	37.4	1.6	1664	9	US-10-091-504-1271
12	37.4	1.6	1664	9	US-10-091-504-1272
13	37.4	1.6	1664	10	US-09-764-869-1271
14	37.4	1.6	1664	10	US-09-764-869-1272
C 15	37.2	1.6	441	9	US-09-796-692-7801
16	37	1.6	5222	10	US-09-993-170-57
17	37	1.6	5696	10	US-09-993-170-60
18	37	1.6	5757	10	US-09-993-170-56
19	36.8	1.6	750	9	US-10-184-644-104

C 20	36.8	1.6	19736	12	US-10-014-502-3	Sequence 3, Appl1
21	36.2	1.5	485	10	US-09-864-761-11473	Sequence 11473, A
C 22	36	1.5	18554	10	US-09-811-825-3	Sequence 3, Appl1
23	35.4	1.5	2671	9	US-09-822-846-129	Sequence 129, Appl
C 24	35.2	1.5	3159	9	US-10-071-766-85	Sequence 85, Appl
25	35.2	1.5	653	9	US-10-184-644-402	Sequence 402, Appl
26	35	1.5	3919	9	US-10-025-384-1125	Sequence 1125, Ap
27	35	1.5	21761	9	US-10-092-154-1680	Sequence 1680, Ap
28	35	1.5	21761	10	US-09-764-847-1680	Sequence 1680, Ap
C 29	35	1.5	46718	9	US-10-274-873-3	Sequence 3, Appl1
30	35	1.5	46718	10	US-09-816-093-3	Sequence 3, Appl1
C 31	34.8	1.5	351	10	US-09-867-701-6593	Sequence 6593, Ap
32	34.8	1.5	671	9	US-10-184-644-346	Sequence 346, Appl
C 33	34.8	1.5	1326	9	US-09-866-570A-53	Sequence 53, Appl
34	34.8	1.5	1326	10	US-09-866-572A-53	Sequence 53, Appl
C 35	34.8	1.5	18596	9	US-09-954-531-124	Sequence 124, Appl
36	34.8	1.5	18596	9	US-09-954-531-348	Sequence 348, Appl
37	34.8	1.5	18596	10	US-09-880-107-1590	Sequence 1590, Ap
C 38	34.8	1.5	18596	10	US-09-967-768A-119	Sequence 119, Appl
C 39	34.6	1.5	378361	9	US-09-901-136-3	Sequence 3, Appl1
40	34.2	1.5	16181	9	US-10-092-154-1426	Sequence 1426, Ap
C 41	34.2	1.5	16181	10	US-09-764-847-1426	Sequence 1426, Ap
42	34	1.4	2139	10	US-09-815-242-6833	Sequence 6833, Ap
C 43	34	1.4	6259	34	US-09-070-927A-129	Sequence 129, Appl
C 44	34	1.4	10877	9	US-10-263-788-1	Sequence 1, Appl1
45	33.8	1.4	3087	10	US-09-815-242-7419	Sequence 7419, Ap

ALIGNMENTS

RESULT 1

US-09-770-445-180/C
Sequence 180, Application US/09770445
Patent No. US2002023281A1
GENERAL INFORMATION:
APPLICANT: Gorlach, Jorn
APPLICANT: An, Yong-Qiang
APPLICANT: Hamilton, Carol M.
APPLICANT: Price, Jennifer L.
APPLICANT: Raines, Tracy M.
APPLICANT: Yu, Yang
APPLICANT: Rameaka, Joshua G.
APPLICANT: Page, Amy
APPLICANT: Matthew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Woessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Krieker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2023US (PARA-012PRV)
CURRENT FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: US 60/178,472
PRIOR FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 180
LENGTH: 1036
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-445-180

Query Match 4.8%; Score 113.2; DB 10; Length 1036;
Best Local Similarity 62.9%; Pred. No. 6.2e-25;
Matches 175; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

```
QY 1570 TTCATGTCGCTCTCGAGCCATCTCTGTGGCCGCTTGTAGTATTCGTGGCA 1629
    |||||
DB 1036 TACATGCTGCAGCGCGCTGAGTGGTGGATGCGCAATCATGTAGTAAAGCAGG 977
QY 1630 AAGCAAACTTTGATCTTGTGAAAAAAGTAGTCTCTCGATGATGATCCTTCGAACT 1689
    |||||
DB 976 AACCAAACTTTGATCTTGTGAAAAAAGTAGTCTCTCGATGATGATCCTTCGAACT 917
QY 1690 GAGTACTATGACATCCGAGCTGCGATTTGTTGTAAGACCCCTTGCATAATGAGAA 1749
    |||||
DB 916 AAGCTCCCGGAGAGGCTTACCTGATCTTATTCGCTGATCCACTAGAGATGGAATC 857
QY 1750 ACTATGCTTAAGATTTGGAATCTCAACAAGTTCATGAGTGTGCTCAACTGC 1809
    |||||
DB 856 AGCTTGCTCAAGTCTGGAACATGATTAAGTTACTGTATTTGTTGTTCAATCTGT 797
QY 1810 CAAGGAGAGATGCTGTGCTGAGACAGCCGCAACCA 1847
DB 796 CAAGGTGCTGTGTTGTCAGAAAGCAAGAAAGAACCA 759
```

RESULT 2

```
US-09-878-574-2803
; Sequence 2803, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 2803
; LENGTH: 356
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-020-Q1-B1-B11
US-09-878-574-2803
```

```
Query Match 3.7%; Score 86.2; DB 10; Length 356;
Best Local Similarity 58.3%; Pred. No. 1e-16;
Matches 151; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
```

```
QY 1476 TCACATGTTCAATGTGCCAAGACAGACAGCTTGTGATGGGGAACCTTCACTCACCCTGACTG 1535
    |||||
DB 91 TCACATGTTCAATGTGCCAAGACAGACAGCTTGTGATGGGGAACCTTCACTCACCCTGACTG 150
QY 1536 GATATGTTCCATCACCACCCCTTGTGATGGGGAACCTTCACTCACCCTGACTG 1595
    |||||
DB 151 GATATGTTCCATCACCACCCCTTGTGATGGGGAACCTTCACTCACCCTGACTG 210
QY 1596 TGGTGGCCGATCATGTTAGTATGTTGTGGAAAGCAATCACTTGTATCTTCTGAAAAA 1655
    |||||
DB 211 GGGATGTCATTAATTAATGATGACAGCCGAGACATGATCTTGTATCTTCTGAAAAA 270
QY 1656 ACTAGTCTCTGATGATGATGATCTTGAAGTGTGATGATGATGATGATGATGATGATGATG 1715
    |||||
DB 271 GCTTGCACACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 330
QY 1716 TGTGTTGTTGAAGACCTT 1734
    |||||
DB 331 TTGCTTATTACTGATCTT 349
```

```
RESULT 3
US-09-294-093B-2999
; Sequence 2999, Application US/09294093B
```

```
; Patent No. US20010051335A1
; GENERAL INFORMATION:
; APPLICANT: Laljudi, Raghunath, V.
; APPLICANT: Ito, Laura, Y.
; APPLICANT: Sherman, Bradley, K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN TASSEL
; FILE REFERENCE: PL-0009 US
; CURRENT APPLICATION NUMBER: US/09/294,093B
; CURRENT FILING DATE: 1999-04-16
; PRIOR APPLICATION NUMBER: 60/082,567
; PRIOR FILING DATE: April 21, 1998
; NUMBER OF SEQ ID NOS: 6207
; SOFTWARE: Perl Program
; SEQ ID NO 2999
; TYPE: DNA
; ORGANISM: Zea mays
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. US20010051335A1 700347006H1
; LOCATION: 190, 193-194, 206, 217, 232, 295
; OTHER INFORMATION: a, t, c, g, or other
US-09-294-093B-2999
```

```
Query Match 3.2%; Score 75.4; DB 10; Length 296;
Best Local Similarity 66.1%; Pred. No. 2.4e-13;
Matches 109; Conservative 0; Mismatches 56; Indels 0; Gaps 0;
```

```
QY 1521 CATCCACCCCTGAGTGGATGTTCCATCCACCCACCCCTTGTGCGCCCTGATGCTGTC 1580
    |||||
DB 1 CATCCACCCCTGAGTGGATGTTCCATCCACCCACCCCTTGTGCGCCCTGATGCTGTC 60
QY 1581 CTTCCAGGACATCTCTGTGCGCCGATCTATGTTAGTATGTTGTGGAAGCATTAATT 1640
    |||||
DB 61 GGTGCGAGGGGTGTGCTGTGCGCATATATGTCACGACAGCCGCGGAGGCATGACTT 120
QY 1641 TGATCTTGTGAAAAAACTAGTGTCTTGTGATGATGATGATGATGATGATGATGATGATG 1685
    |||||
DB 121 CATCTGCTCAGAAAGACCTGCTGCTCCGACGATGATGCTGCG 165
```

RESULT 4

```
US-09-878-574-1840
; Sequence 1840, Application US/09878574
; Patent No. US20020110548A1
; GENERAL INFORMATION:
; APPLICANT: Byrum, Joseph R.
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Thompson, Michael D.
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; FILE OF INVENTION: Plants
; FILE REFERENCE: 38-21(15401)B
; CURRENT APPLICATION NUMBER: US/09/878,574
; CURRENT FILING DATE: 2001-12-21
; PRIOR APPLICATION NUMBER: 09/333,535
; PRIOR FILING DATE: 1999-06-14
; NUMBER OF SEQ ID NOS: 15775
; SEQ ID NO 1840
; LENGTH: 377
; TYPE: DNA
; ORGANISM: Glycine max
; OTHER INFORMATION: Clone ID: LIB3028-032-Q1-B1-E5
US-09-878-574-1840
```

```
Query Match 1.9%; Score 45.8; DB 10; Length 377;
Best Local Similarity 55.3%; Pred. No. 0.00068;
Matches 89; Conservative 0; Mismatches 72; Indels 0; Gaps 0;
```

```
QY 192 TGGTCTTGTGTTGATGACGCGTGGACCTGATTAAGCAGATGTTGTTGATTTGG 251
    |||||
DB 217 TGGGATTCATTTGGGTTCAATCCATCACAAGGTACCCGACAGTCTTCCCAATTGG 276
```



```
QY 497 TTGATGCATCGTTCGAGATGTGTATCTTCATGCTGCTGATGATCCGTTTGCACTTG 556
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641 TTAATCCCCACTTCCAAATGATATTATGTAATGCGCATTTGGTACTGCAATATGCAATGG 700
QY 557 TTAAAG 562
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 ATTAATG 706

RESULT 8
US-09-866-572A-19
; Sequence 19, Application US/09866572A
; Patent No. US20020138859A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/866,572A
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 09/457,046
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Taxus cuspidata
US-09-866-572A-19

Query Match 1.6%; Score 38; DB 10; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.37;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 317 TTGGTGAATGTGGGGATCTTGATCGAGACTGAGATTGATGCTTGAGAGATCGAG 376
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 TTGTTTAATGTGTGACCATGAAATATATGAACAATGTGTTAGGAGAAATGTAATG 520
QY 377 ATTCGTGACCGATGATGTTTCTTCCTTCGATCGTGAAGGACCGTTCCGAACCTCGA 436
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 ATTTGTTGTCGTCCTTTGAAGTAGTGAGCATTTGGATAGCAAGCAAAAGGCTC 580
QY 437 TTCAGCCTGGGATGATGACTTTGTCATGCTTGTGTCCAGAGTGGTTCGCAAGCTTG 496
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 TTCAAAATTCACATGACTGAGATGAGACCTTCCTTTGCGAGGATTTGAGGAAATTTAT 640
QY 497 TTGATGCATCGTTCGAGATGTGTATCTTCATGCTGCTGATGATCCGTTTGCACTTG 556
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641 TTAATCCCCACTTCCAAATGATATTATGTAATGCGCATTTGGTACTGCAATATGCAATGG 700
QY 557 TTAAAG 562
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 ATTAATG 706

RESULT 9
US-09-866-572A-21
; Sequence 21, Application US/09866572A
; Patent No. US20020138859A1
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney et al.
; TITLE OF INVENTION: Transacylases of the Paclitaxel Biosynthetic Pathway
; FILE REFERENCE: 53679
; CURRENT APPLICATION NUMBER: US/09/866,572A
; CURRENT FILING DATE: 2001-05-25
; PRIOR APPLICATION NUMBER: 09/457,046
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 74
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 911
; TYPE: DNA
; ORGANISM: Taxus cuspidata
US-09-866-572A-21
```

```
Query Match 1.6%; Score 38; DB 10; Length 911;
Best Local Similarity 47.2%; Pred. No. 0.37;
Matches 116; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

QY 317 TTGGTGAATGTGGGGATCTTGATCGAGACTGAGATTGATGCTTGAGAGATCGAG 376
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 461 TTGTTTAATGTGTGACCATGAAATATATGAACAATGTGTTAGGAGAAATGTAATG 520
QY 377 ATTCGTGACCGATGATGTTTCTTCCTTCGATCGTGAAGGACCGTTCCGAACCTCGA 436
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 521 ATTTGTTGTCGTCCTTTGAAGTAGTGAGCATTTGGATAGCAAGCAAAAGGCTC 580
QY 437 TTCAGCCTGGGATGATGACTTTGTCATGCTTGTGTCCAGAGTGGTTCGCGAAAGTTG 496
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 581 TTCAAAATTCACATGACTGAGATGAGACCTTCCTTTGCGAGGATTTGAGGAAATTTAT 640
QY 497 TTGATGCATCGTTCGAGATGTGTATCTTCATGCTGCTGATGATCCGTTTGCACTTG 556
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 641 TTAATCCCCACTTCCAAATGATATTATGTAATGCGCATTTGGTACTGCAATATGCAATGG 700
QY 557 TTAAAG 562
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 701 ATTAATG 706

RESULT 10
US-09-820-003A-3
; Sequence 3, Application US/09820003A
; Patent No. US20020142382A1
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Gennady et al.
; TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,
; TITLE OF INVENTION: NOCULEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE
; TITLE OF INVENTION: PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL001196
; CURRENT APPLICATION NUMBER: US/09/820,003A
; CURRENT FILING DATE: 2001-03-29
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 46050
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(46050)
; OTHER INFORMATION: n = A,T,C or G
US-09-820-003A-3

Query Match 1.6%; Score 37.6; DB 10; Length 46050;
Best Local Similarity 52.6%; Pred. No. 7.7;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1809 CCNAGGAGGATGTGTGTCGAGACCGCGCAACCAATGCTTTTCACAACTACTCAAA 1868
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30070 CCCAGGAGGGGAGGTGTGTGTCGAGTCGAGATGCGACCAATTTGCATCCAGCGTGGCAAC 30129
QY 1869 ACAGTGCATCTCAAAACTTAACCCAAAGACATAGATGGCACATGGAGGAAAACCTTAT 1928
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30130 AAGAGCGAAACTCCATCTCAAAAAAAGGAGAGGATATATACAAAT 30189
QY 1929 CTCTATTGAAGGCGTTAAACCTTTGCGCTTACCT 1964
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 30190 TGCATTAATATGCAAAAAAAGGCTCAACTT 30225

RESULT 11
US-10-091-504-1271
; Sequence 1271, Application US/10091504
; Publication No. US20030059908A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
```

FILE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091.504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1271
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1271

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88; Length 1664;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Oy 1782 CACTGAGTGGTGGTGCATTCACCTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
Db 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGCGAGCGAGTTCAGTGCAGCCGAGATC 1046
Oy 1842 CAACCAATGCTTTTCACATCTCACTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
Db 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAGAGGAACCTCATCTCAAAAAAAAAAAAA 1106
Oy 1902 AGAATGCGACAGTGGAGAAACCCATCTCT 1932
Db 1107 AGAAAAAGAAAAAGAAAGAAACTTTCCCT 1137

RESULT 12
US-10-091-504-1272
Sequence 1272, Application US/10091504
Publication No. US20030059908A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007C1
CURRENT APPLICATION NUMBER: US/10/091.504
CURRENT FILING DATE: 2002-03-07
NUMBER OF SEQ ID NOS: 2442
Prior Application removed - See File Wrapper or Palm
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1272
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-10-091-504-1272

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88; Length 1664;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Oy 1782 CACTGAGTGGTGGTGCATTCACCTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
Db 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGCGAGCGAGTTCAGTGCAGCCGAGATC 1046
Oy 1842 CAACCAATGCTTTTCACATCTCACTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
Db 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAGAGGAACCTCATCTCAAAAAAAAAAAAA 1106
Oy 1902 AGAATGCGACAGTGGAGAAACCCATCTCT 1932
Db 1107 AGAAAAAGAAAAAGAAAGAAACTTTCCCT 1137

RESULT 13
US-09-764-869-1271
Sequence 1271, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764, 869
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1271
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1271

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88; Length 1664;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Oy 1782 CACTGAGTGGTGGTGCATTCACCTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
Db 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGCGAGCGAGTTCAGTGCAGCCGAGATC 1046
Oy 1842 CAACCAATGCTTTTCACATCTCACTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
Db 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAGAGGAACCTCATCTCAAAAAAAAAAAAA 1106
Oy 1902 AGAATGCGACAGTGGAGAAACCCATCTCT 1932
Db 1107 AGAAAAAGAAAAAGAAAGAAACTTTCCCT 1137

RESULT 14
US-09-764-869-1272
Sequence 1272, Application US/09764869
Patent No. US20020061521A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC007
CURRENT APPLICATION NUMBER: US/09/764, 869
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PALM or file wrapper
NUMBER OF SEQ ID NOS: 2442
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1272
LENGTH: 1664
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-869-1272

Query Match
Best Local Similarity 53.0%; Pred. No. 0.88; Length 1664;
Matches 80; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

Oy 1782 CACTGAGTGGTGGTGCATTCACCTGCCAAGAGAGAGAGTGTCTGTGAGACACGCCG 1841
Db 987 CACTCGGAGCGCTGAGCAGAGAGTGAACCCGCGAGCGAGTTCAGTGCAGCCGAGATC 1046
Oy 1842 CAACCAATGCTTTTCACATCTCACTCAAAAGAGTGCATCCAAACTAACCCAAAGACAT 1901
Db 1047 GCGCCAGTGCACCTTCAAGCTGGGCAACAGAGGAACCTCATCTCAAAAAAAAAAAAA 1106
Oy 1902 AGAATGCGACAGTGGAGAAACCCATCTCT 1932
Db 1107 AGAAAAAGAAAAAGAAAGAAACTTTCCCT 1137

RESULT 15
US-09-796-692-7801/C
Sequence 7801, Application US/09796692
Publication No. US20020198362A1
GENERAL INFORMATION:
APPLICANT: Gaiger, Alexander
APPLICANT: Algate, Paul A.
APPLICANT: Mannion, Jane

```
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE OF INVENTION: HEMATOLOGICAL MALIGNANCIES
; FILE REFERENCE: 2077.001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 7801
; LENGTH: 441
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-796-692-7801

Query Match      1.6%; Score 37.2; DB 9; Length 441;
Best Local Similarity 57.9%; Pred. No. 0.4;
Matches 66; Conservative 0; Mismatches 48; Indels 0; Gaps 0;

QY 1809 CCAAGAGAGAGATGCTGCTGAGACACGCCCAACCAATGCTTTTCACATATCTCAAA 1868
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 328 CCTGGAGGCGGAGGTGTGTGAGCGGAGATGCGACCATTTGCATTCACGCTGGGCAAC 269
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1869 ACGAGTGACATCCAAACTAACCCAAAGACATAGATGCGACATGGAGAAAA 1922
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 268 AAGAGCGAACTCGCTCTCAAAAAAAAAAGAGAAAAAGAAAAAGAAAAA 215
```

Search completed: April 3, 2003, 21:26:43
Job time : 398 secs

GenCore version 5.1.4-P5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 3.41463 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-3

Perfect score: 80
Sequence: 1 YDODMWWVQVWP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Issued Patents-AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilltest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	80	100.0	14	US-08-846-234-3	Sequence 3, Appl1
2	80	100.0	784	US-08-846-234-5	Sequence 5, Appl1
3	38	47.5	404	US-09-302-769-21	Sequence 21, Appl1
4	38	47.5	527	US-09-659-166-2	Sequence 2, Appl1
5	38	47.5	528	US-08-802-466-2	Sequence 2, Appl1
6	38	47.5	528	US-09-350-484-2	Sequence 2, Appl1
7	38	47.5	1829	US-09-157-420-1	Sequence 1, Appl1
8	37	46.2	156	US-09-064-922-2	Sequence 2, Appl1
9	37	46.2	179	US-09-064-922-5	Sequence 5, Appl1
10	37	46.2	399	US-09-491-362-7	Sequence 7, Appl1
11	37	46.2	399	US-08-874-562-7	Sequence 16, Appl1
12	37	46.2	452	US-08-764-870-15	Sequence 16, Appl1
13	37	46.2	452	US-08-980-115-16	Sequence 16, Appl1
14	37	46.2	475	US-09-491-362-2	Sequence 2, Appl1
15	37	46.2	475	US-09-874-562-2	Sequence 2, Appl1
16	37	46.2	477	US-09-449-335-2	Sequence 6, Appl1
17	37	46.2	477	US-09-449-335-6	Sequence 10, Appl1
18	37	46.2	477	US-09-480-921B-10	Sequence 29, Appl1
19	37	46.2	477	US-09-480-921B-29	Sequence 11, Appl1
20	37	46.2	918	US-09-041-886-11	Sequence 265, Appl1
21	36	45.0	229	US-09-199-637A-265	Sequence 4, Appl1
22	36	45.0	400	US-09-184-001-4	Sequence 4, Appl1
23	36	45.0	404	US-09-184-001-2	Sequence 4, Appl1
24	36	45.0	3170	US-09-036-987A-4	Sequence 85, Appl1
25	35	43.8	18	US-09-370-700-4	Sequence 3298, Appl1
26	35	43.8	18	US-08-484-192-85	
27	35	43.8	347	US-09-134-001C-3298	

28	35	43.8	473	1	US-08-597-236-13	Sequence 13, Appl1
29	35	43.8	473	1	US-08-746-682A-13	Sequence 13, Appl1
30	35	43.8	563	4	US-09-718-841-2	Sequence 2, Appl1
31	35	43.8	563	4	US-09-718-810-2	Sequence 2, Appl1
32	35	43.8	589	2	US-07-668-646-6	Sequence 6, Appl1
33	35	43.8	589	2	US-08-429-998-6	Sequence 6, Appl1
34	35	43.8	589	2	US-08-431-333-6	Sequence 6, Appl1
35	35	43.8	589	5	PCT-US91-02321-6	Sequence 6, Appl1
36	35	43.8	1385	1	US-07-876-280-2	Sequence 2, Appl1
37	35	43.8	1385	1	US-07-675-772-2	Sequence 2, Appl1
38	35	43.8	1385	1	US-08-063-170-2	Sequence 2, Appl1
39	35	43.8	1385	1	US-08-158-232-2	Sequence 2, Appl1
40	35	43.8	1385	1	US-08-304-626-2	Sequence 2, Appl1
41	35	43.8	1385	1	US-08-316-301A-2	Sequence 2, Appl1
42	35	43.8	1385	2	US-08-611-928-2	Sequence 2, Appl1
43	35	43.8	1385	3	US-09-173-891-2	Sequence 2, Appl1
44	35	43.8	1385	4	US-09-076-137-2	Sequence 2, Appl1
45	35	43.8	1385	5	PCT-US92-03624-2	Sequence 2, Appl1

ALIGNMENTS

```
RESULT 1
US-08-846-234-3
; Sequence 3, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MATER & NEUSTADT, P. C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24 618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 14 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; US-08-846-234-3

Query Match      100.0%; Score 80; DB 4; Length 14;
Best Local Similarity 100.0%; Pred. No. 5.4e+08;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 YDODMWWVQVWP 14
Db      1 YDODMWWVQVWP 14
```

```
RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-846-234-5

Query Match          100.0%; Score 80; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 5.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 YDDQDMVVVQVMP 14
        |||||
Db       756 YDDQDMVVVQVMP 769

RESULT 3
US-09-302-769-21
; Sequence 21, Application US/09302769
; Patent No. 6323317
; GENERAL INFORMATION:
; APPLICANT: HILTON, Douglas J
; APPLICANT: ALEXANDER, Warren S
; APPLICANT: VINEY, Elizabeth M
; APPLICANT: WILSON, Tracey A
; APPLICANT: RICHARDSON, Rachael T
; APPLICANT: STARR, Robyn
; APPLICANT: NICHOLSON, Sandra E
; APPLICANT: METCALF, Donald
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC AGENTS
; FILE REFERENCE: 109762
; CURRENT APPLICATION NUMBER: US/09/302,769
; CURRENT FILING DATE: 1999-04-30
; PRIOR APPLICATION NUMBER: 08/962,560
; PRIOR FILING DATE: 1997-10-31
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 21
; LENGTH: 404
```

```
; TYPE: PRT
; ORGANISM: Human
US-09-302-769-21

Query Match          47.5%; Score 38; DB 4; Length 404;
Best Local Similarity 50.0%; Pred. No. 58;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY      1 YDDQDMVVVQVMP 14
        : | || |||
Db       43 WSGCHGVKIVMP 56

RESULT 4
US-09-659-166-2
; Sequence 2, Application US/09659166
; Patent No. 6355465
; GENERAL INFORMATION:
; APPLICANT: GREASY, CAREYNA LEE
; APPLICANT: TESTA, TANIA TAMSON
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP-30218
; CURRENT APPLICATION NUMBER: US/09/659,166
; CURRENT FILING DATE: 2000-09-11
; PRIOR APPLICATION NUMBER: UK 9921505.5
; PRIOR FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 527
; TYPE: PRT
; ORGANISM: rattus
; US-09-659-166-2

Query Match          47.5%; Score 38; DB 4; Length 527;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      1 YDDQDMVVVQVMP 12
        || || |||
Db       129 YDDQGSYVQVP 140

RESULT 5
US-08-802-466-2
; Sequence 2, Application US/08802466
; Patent No. 5972606
; GENERAL INFORMATION:
; APPLICANT: Creasy, et al.
; TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION NUMBER: US/08/802,466
; FILING DATE: 19 February 1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: William T. Han
; REGISTRATION NUMBER: 34,344
```


REFERENCE/DOCKET NUMBER: GH50002
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-802-466-2

Query Match 47.5%; Score 38; DB 2; Length 528;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 YDDQDMVVQVP 12
11111111
Db 130 YDDQDGSYVQVP 141

RESULT 6
US-09-350-484-2
Sequence 2, Application US/09350484
Patent No. 6159716
GENERAL INFORMATION:
APPLICANT: Creasy, et al.
TITLE OF INVENTION: HUMAN PROTEIN KINASE HOACF72
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/350,484
FILING DATE: 09-Jul-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/802,466
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: William T. Han
REGISTRATION NUMBER: 34,344
REFERENCE/DOCKET NUMBER: GH50002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-5219
TELEFAX: 610-270-4026
TELEX: <Unknown>
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 528 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-350-484-2

Query Match 47.5%; Score 38; DB 4; Length 528;
Best Local Similarity 66.7%; Pred. No. 79;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 1 YDDQDMVVQVP 12

Db 130 YDDQDGSYVQVP 141

RESULT 7
US-09-157-420-1
Sequence 1, Application US/09157420
Patent No. 6180760
GENERAL INFORMATION:
APPLICANT: TAKAI, Yoshimi
APPLICANT: NAKANISHI, Hiroyuki
APPLICANT: MANDAI, Kenji
APPLICANT: WADA, Manabu
APPLICANT: OBAISHI, Hiroshi
TITLE OF INVENTION: ACTIN FILAMENT-BINDING PROTEIN "L-AFADIN"
FILE REFERENCE: 98-1042/LC(WMC)/653
CURRENT APPLICATION NUMBER: US/09/157,420
CURRENT FILING DATE: 1998-09-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 1829
TYPE: PRT
ORGANISM: rat
US-09-157-420-1

Query Match 47.5%; Score 38; DB 4; Length 1829;
Best Local Similarity 58.3%; Pred. No. 3.4e+02;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 3 QDMVVQVPWP 14
11111111
Db 1246 QDMAPVQNPW 1257

RESULT 8
US-09-064-922-2
Sequence 2, Application US/09064922
Patent No. 622095
GENERAL INFORMATION:
APPLICANT: Callis, Judy
APPLICANT: Worley, Cathy K.
TITLE OF INVENTION: Sequences from Auxin-Induced Gene
TITLE OF INVENTION: Products Targeting Fusion Proteins for Degradation
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/064,922
FILING DATE: 22-Apr-1998
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Hyman, Laurence J.
REGISTRATION NUMBER: 35,551
REFERENCE/DOCKET NUMBER: 023070-08540005
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 156 amino acids
TYPE: amino acid
TOPOLOGY: linear


```

; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,870
; FILING DATE: 13-DEC-1996
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,540
; FILING DATE: 13-DEC-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,543
; FILING DATE: 13-DEC-1995
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 60/008,606
; FILING DATE: 14-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Nakamura, Jackie N
; REGISTRATION NUMBER: 35,966
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)843-5000
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 452 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-764-870-16

Query Match      46.2%; Score 37; DB 4; Length 452;
Best Local Similarity 60.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 DOMVVVQVPM 13
    1111111111
Db 265 DOMAVIQYSM 274

RESULT 13
US-08-980-115-16
; Sequence 16, Application US/08980115
; GENERAL INFORMATION:
; APPLICANT: Scanlan, Thomas S.
; APPLICANT: Baxter, John D.
; APPLICANT: Fletcherick, Robert J.
; APPLICANT: Wagner, Richard L.
; APPLICANT: Kushner, Peter J.
; APPLICANT: Apilletl, James W.
; APPLICANT: West, Brian L.
; APPLICANT: Shlau, Andrew K.
; TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
; FILE REFERENCE: UCAL-246/0205
; CURRENT APPLICATION NUMBER: US/08/980,115
; CURRENT FILING DATE: 1997-11-26
; EARLIER APPLICATION NUMBER: 08/764,870
; EARLIER FILING DATE: 1996-12-13
; EARLIER APPLICATION NUMBER: 60/008,606
; EARLIER FILING DATE: 1995-12-14
; EARLIER APPLICATION NUMBER: 60/008,543
; EARLIER FILING DATE: 1995-12-13
; EARLIER APPLICATION NUMBER: 60/008,540
; EARLIER FILING DATE: 1995-12-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 16
; LENGTH: 452
```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (184)..(437)
; OTHER INFORMATION: minimal ligand binding domain
; US-08-980-115-16

Query Match      46.2%; Score 37; DB 4; Length 452;
Best Local Similarity 60.0%; Pred. No. 99;
Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 4 DOMVVVQVPM 13
    1111111111
Db 265 DOMAVIQYSM 274

RESULT 14
US-09-491-362-2
; Sequence 2, Application US/09491362
; Patent No. 6281017
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: 1-DEOXY-D-XYLOULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
; FILE REFERENCE: WSUR14977
; CURRENT APPLICATION NUMBER: US/09/491,362
; CURRENT FILING DATE: 2000-01-26
; EARLIER APPLICATION NUMBER: 60/118,349
; EARLIER FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Mentha piperita
; US-09-491-362-2

Query Match      46.2%; Score 37; DB 4; Length 475;
Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 QDOMVVQVPM 14
    1111111111
Db 331 QDSSVLAQLGMP 342

RESULT 15
US-09-874-562-2
; Sequence 2, Application US/09874562
; Patent No. 6420159
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B
; APPLICANT: Lange, Bernd M
; TITLE OF INVENTION: 1-DEOXY-D-XYLOULOSE-5-PHOSPHATE REDUCTOISOMERASE, AND
; FILE REFERENCE: WSUR17549
; CURRENT APPLICATION NUMBER: US/09/874,562
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/491,362
; PRIOR FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 60/118,349
; PRIOR FILING DATE: 1999-02-03
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 475
; TYPE: PRT
; ORGANISM: Mentha piperita
; US-09-874-562-2

Query Match      46.2%; Score 37; DB 4; Length 475;
```

Best Local Similarity 50.0%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 3 ODOMVVOYVWP 14
|||:|:|
Db 331 QDSSVLAQLGWP 342

Search completed: March 27, 2003, 10:07:35
Job time : 4.41463 secs

CC Raffinose synthase forms raffinose from sucrose and galactinol, has
 CC an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees
 CC C, has a molecular weight of 75 to 95 kDa by gel filtration or 90
 CC to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is
 CC inhibited by Iodacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 14 AA:

Query Match 100.0%; Score 80; DB 19; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVQVWP 14
 DB 1 YDQDMVVQVWP 14

RESULT 2

AA17424
 ID AAY17424 standard; peptide; 14 AA.

XX AAY17424;

DT 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase peptide SEQ ID NO:3.

XX Cucurbitaceae; sucrose; galactinol.

XX Cucumis sativus.

XX JP1123080-A.

XX 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1999-340516/29.

PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

PS Example 2; Page 22; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents a raffinose
 CC synthase peptide from cucumber.

XX Sequence 14 AA:

Query Match 100.0%; Score 80; DB 20; Length 14;
 Best Local Similarity 100.0%; Pred. No. 1.7e-07;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVQVWP 14
 DB 1 YDQDMVVQVWP 14

RESULT 3

AAW53570
 ID AAW53570 standard; Protein; 784 AA.

XX AAW53570;

DT 06-JUL-1998 (first entry)

XX Cucumber raffinose synthase.

XX Cucurbitaceae; raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

XX JP10084973-A.

XX 07-APR-1998.

PF 28-APR-1997; 97JP-0111124.

PR 26-JUL-1996; 96JP-0198079.

PR 26-APR-1996; 96JP-0107682.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1998-264858/24.

XX N-PSDB; AAW22250.

PT Raffinose synthase gene - useful for preparation of raffinose in
 PT transformed plant

PS Claim 3; Pages 17-20; 26pp; Japanese.

CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC Iodacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA:

Query Match 100.0%; Score 80; DB 19; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVQVWP 14
 DB 756 YDQDMVVQVWP 769

RESULT 4

AA17417
 ID AAY17417 standard; Protein; 784 AA.

XX AAY17417;

DT 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase.

XX Cucurbitaceae; sucrose; galactinol.

XX Cucumis sativus.

XX JP1123080-A.

XX 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

XX WPI; 1999-340516/29.

XX N-PSDB; AAX61238.

PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

XX Claim 2; Page 25-27; 37pp; Japanese.
 PS
 CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents raffinose
 CC synthase from cucumber.
 XX

SQ Sequence 784 AA;

Query Match 100.0%; Score 80; DB 20; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.5e-05;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 YDQDMVVVQVWMP 14
 DB 756 YDQDMVVVQVWMP 769

RESULT 5

AAB98659
 ID AAB98659 standard; protein; 780 AA.

AC AAB98659;

DT 17-AUG-2001 (first entry)

DE Soybean protein: SEQ ID 1.

KW Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;
 KM plant; soybean.

XX Glycine max.

PN JP2001078783-A.

PD 27-MAR-2001.

PF 03-JUL-2000; 2000JP-0200571.

PR 09-JUL-1999; 99JP-0196036.

PA (SUMO) SUMITOMO CHEM CO LTD.

DR WPI: 2001-313373/33.

DR N-PSDB: AAH27438.

PT Novel mutant protein of raffinose synthase is useful for reducing the
 PF raffinose oligosaccharide content in a plant body -

PS Disclosure; Page 18-20; 30pp; Japanese.

CC The present invention relates to a mutant protein of raffinose synthase
 CC in which at least one aromatic amino acid present at the position of
 CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
 CC mutant protein can be used for reducing the raffinose oligosaccharide
 CC content in a plant body. The present protein from soybean, was used in
 CC the present invention.
 XX

SQ Sequence 780 AA;

Query Match 62.5%; Score 50; DB 22; Length 780;
 Best Local Similarity 66.7%; Pred. No. 2.9;

Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 OQDMVVVQVWMP 14
 DB 757 EDKMLRVQVWMP 768

RESULT 6

AAW57887
 ID AAW57887 standard; protein; 781 AA.

AC AAW57887;

DT 23-SEP-1998 (first entry)

DE Soybean raffinose synthetase.

KW Raffinose synthetase; metabolism modification; food additive;
 KM gastrointestinal flora; soybean.

XX Glycine max.

PN EP849359-A2.

PD 24-JUN-1998.

PF 18-DEC-1997; 97EP-0122417.

PR 18-DEC-1996; 96JP-0338673.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Oeda K, Wantanabe E;

DR WPI: 1998-324670/29.

DR N-PSDB: AAV40801.

PT New nucleic acid molecule encoding plant raffinose synthetase -
 PF capable of producing raffinose, used as food additives with
 PT beneficial effects on gastrointestinal flora

PS Claim 1; Page 31-34; 44pp; English.

CC This sequence represents the soybean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.
 XX

SQ Sequence 781 AA;

Query Match 62.5%; Score 50; DB 19; Length 781;
 Best Local Similarity 66.7%; Pred. No. 2.9;

Matches 8; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 3 OQDMVVVQVWMP 14
 DB 758 EDKMLRVQVWMP 769

RESULT 7

AAV30143
 ID AAV30143 standard; protein; 781 AA.

AC AAV30143;

DT 26-OCT-1999 (first entry)

DE Amino acid sequence of a raffinose synthase protein.

KW Raffinose synthase; plant; sucrose; raffinose.

XX Glycine max.

PN JP11215984-A.

PD 10-AUG-1999.


```
XX XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX DR WPI: 2001-616774/71.
XX DR N-PSDB: AAS59786.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX PS Example 1: SEQ ID NO 28413; 1069pp; English.
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 98 AA:
Query Match 57.5%; Score 46; DB 22; Length 98;
Best Local Similarity 77.8%; Pred. NO. 1.5;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 6 MVTVOVPWP 14
| | | | |
DB 9 MVTVOIPWP 17
| | | | |
RESULT 11
AAU49825
ID AAU49825 standard; Protein: 148 AA.
XX AC AAU49825;
XX AC
XX DT 27-FEB-2002 (first entry)
XX DE Propionibacterium acnes immunogenic protein #10721.
```

```
XX XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW dermatological; osteopathic; neuroprotectant.
XX OS Propionibacterium acnes.
XX PN WO200181581-A2.
XX PD 01-NOV-2001.
XX PF 20-APR-2001; 2001WO-US12865.
XX PR 21-APR-2000; 2000US-199047P.
XX PR 02-JUN-2000; 2000US-208841P.
XX PR 07-JUL-2000; 2000US-216747P.
XX PA (CORI-) CORIXA CORP.
XX PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhatia A;
XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX DR WPI: 2001-616774/71.
XX DR N-PSDB: AAS59545.
XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
XX PT vaccinating against and diagnosing infections, especially useful for
XX PT treating acne vulgaris -
XX PS Example 1: SEQ ID NO 11020; 1069pp; English.
XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
XX CC polypeptides. The proteins and their associated DNA sequences are used in
XX CC the treatment, prevention and diagnosis of medical conditions caused by
XX CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
XX CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
XX CC P. acnes is also involved in infections of bone, joints and the central
XX CC nervous system, however it is particularly involved in the inflammatory
XX CC lesions associated with acne vulgaris. A method for detecting the
XX CC presence or absence of P. acnes in a patient comprises contacting a
XX CC sample with a binding agent that binds to the proteins of the invention
XX CC and determining the amount of bound protein in the sample. The
XX CC polypeptides may be used as antigens in the production of antibodies
XX CC specific for P. acnes proteins. These antibodies can be used to
XX CC downregulate expression and activity of P. acnes polypeptides and
XX CC therefore treat P. acnes infections. The antibodies may also be used as
XX CC diagnostic agents for determining P. acnes presence, for example, by
XX CC enzyme linked immunosorbent assay (ELISA).
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 148 AA:
Query Match 57.5%; Score 46; DB 22; Length 148;
Best Local Similarity 77.8%; Pred. NO. 2.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 6 MVTVOVPWP 14
| | | | |
DB 9 MVTVOIPWP 17
| | | | |
RESULT 12
AAG53157
ID AAG53157 standard; Protein: 484 AA.
XX AC AAG53157;
XX AC
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 67650.
```

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 07-MAY-1999; 990S-0132487.
PR 11-MAY-1999; 990S-0132863.
PR 14-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
PR 01-JUN-1999; 990S-0137222.
PR 03-JUN-1999; 990S-0137528.
PR 04-JUN-1999; 990S-0137502.
PR 07-JUN-1999; 990S-0137724.
PR 08-JUN-1999; 990S-0138094.
PR 10-JUN-1999; 990S-0138540.
PR 10-JUN-1999; 990S-0138847.
PR 14-JUN-1999; 990S-0139119.
PR 16-JUN-1999; 990S-0139452.
PR 16-JUN-1999; 990S-0139453.
PR 17-JUN-1999; 990S-0139459.
PR 17-JUN-1999; 990S-0139454.
PR 18-JUN-1999; 990S-0139455.
PR 18-JUN-1999; 990S-0139456.
PR 18-JUN-1999; 990S-0139457.
PR 18-JUN-1999; 990S-0139458.
PR 18-JUN-1999; 990S-0139459.
PR 18-JUN-1999; 990S-0139460.
PR 18-JUN-1999; 990S-0139461.
PR 18-JUN-1999; 990S-0139462.
PR 18-JUN-1999; 990S-0139463.
PR 18-JUN-1999; 990S-0139750.
PR 18-JUN-1999; 990S-0139763.
PR 21-JUN-1999; 990S-0139817.
PR 22-JUN-1999; 990S-0139899.

PR 23-JUN-1999; 990S-0140353.
PR 23-JUN-1999; 990S-0140354.
PR 24-JUN-1999; 990S-0140695.
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 02-JUL-1999; 990S-0142154.
PR 06-JUL-1999; 990S-0142055.
PR 08-JUL-1999; 990S-0142390.
PR 09-JUL-1999; 990S-0142803.
PR 12-JUL-1999; 990S-0142920.
PR 13-JUL-1999; 990S-0142977.
PR 14-JUL-1999; 990S-0143542.
PR 15-JUL-1999; 990S-0143624.
PR 16-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 05-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 06-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.


```

PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

```

```

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0155658.
PR 29-SEP-1999; 99US-0155596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 56.2%; Score 45; DB 21; Length 563;
 Best Local Similarity 53.8%; Pred No. 15;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 DDDQWVWQVMP 14
 Db 536 DEDMLVGDPPM 548

```

RESULT 14
AAV32075
ID AAV32075 standard; Protein; 572 AA.
XX
AC AAV32075;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rapeseed raffinose synthase.
XX
KW Raffinose synthase; rapeseed; transgenic plant.
XX
OS Brassica napus.
XX
FH Key Location/Qualifiers
FT Misc-difference 129 /note= "encoded by GGY"
FT Misc-difference 132 /note= "encoded by GGY"
FT Misc-difference 143 /note= "encoded by GSW"
FT Misc-difference 144 /note= "encoded by CCS"
FT Misc-difference 144 /note= "encoded by TCR"
FT

```


GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 4.0122 Seconds
(without alignments)
335.448 Million cell updates/sec

Title: US-09-675-208-3
Perfect score: 80
Sequence: 1 YDDQWVYVQVWMP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: pir3:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	55.0	164	2 S58501	auxin-induced prot
2	44	55.0	225	2 H71407	auxin-induced prote
3	44	55.0	229	2 H86173	hypothetical prote
4	44	55.0	234	2 C85159	IAA7 like protein
5	44	55.0	619	2 C86467	hypothetical prote
6	44	55.0	719	2 T47727	hypothetical prote
7	43	53.8	418	2 G97360	intracellular PHB
8	43	53.8	418	2 AH2578	intracellular PHB
9	43	53.8	423	2 T52251	5'-adenylyl sulfate
10	43	53.8	863	1 S51789	VLDL receptor prec
11	42	52.5	500	2 D86274	hypothetical prote
12	42	52.5	529	2 T42584	tegument protein 4
13	42	52.5	600	2 C86468	probable auxin res
14	42	52.5	767	2 C86476	protein F1504.37 (
15	41	51.2	530	1 W2BED4	gene 40 protein -
16	40	50.0	189	2 S58493	auxin-induced prot
17	40	50.0	189	2 E95659	probable IAA6 prot
18	40	50.0	346	2 AG2094	hypothetical prote
19	40	50.0	448	2 P95036	pts system, IIC co
20	40	50.0	448	2 B92907	phosphotransferase
21	40	50.0	653	2 S75763	acetate-CoA ligase
22	40	50.0	1159	2 T43461	probable phosphodi
23	39	48.8	206	2 S22400	insecticynin A -
24	39	48.8	281	2 A65219	phn protein - Esc
25	39	48.8	281	2 A91264	phosphonate metabo
26	39	48.8	281	2 F86104	phosphonate metabo
27	39	48.8	305	2 T08963	phytochrome-associ
28	39	48.8	384	2 G82670	general secretory
29	39	48.8	848	2 JG0194	androgen receptor

30	38	47.5	162	2 C82765	conserved hypothet
31	38	47.5	190	2 C64950	yeam protein - Esc
32	38	47.5	190	2 A09952	hypothetical prote
33	38	47.5	190	2 E85800	hypothetical prote
34	38	47.5	305	2 A25010	pectinesterase (EC
35	38	47.5	390	2 A70656	glutamate-ammonia
36	38	47.5	432	2 S22527	glutamate-beta-1,3-
37	38	47.5	473	2 E84471	conserved hypothet
38	38	47.5	524	2 D82220	probable membrane
39	38	47.5	569	2 C86934	hypothetical prote
40	38	47.5	624	2 B96536	hypothetical prote
41	38	47.5	722	2 T30995	probable reverse t
42	38	47.5	1206	2 S72620	s-afadin - rat
43	38	47.5	1663	2 T42092	phosphodiesterase
44	38	47.5	1687	2 T30244	exoribonuclease, v
45	38	47.5	1706	2 T30175	

ALIGNMENTS

RESULT 1
S58501
auxin-induced protein IAA14 - Arabidopsis thaliana (fragment)
N:Alternate names: indoleacetic acid-inducible protein IAA14
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 22-Nov-1996 #sequence_revision 22-Nov-1996 #text_change 20-Aug-1999
C:Accession: S58501; S71408
R:Abel, S.; Nguyen, M.D.; Theologis, A.
J. Mol. Biol. 251, 533-549, 1995
A>Title: The ps-IAA4/5-like family of early auxin-inducible mRNAs in Arabidopsis thal
A:Reference number: S58491; MUID:95387593; PMID:7658471
A:Accession: S58501
A>Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-164 <ABE>
R:Theologis, A.
submitted to the EMBL Data Library, December 1994
A:Reference number: S71406
A:Accession: S71408
A:Molecule type: DNA
A:Residues: 1-109, 'S', 111-164 <THE>
A:Cross-references: EMBL:U18416; NID:g972930; PIDN:AAC49055.1; PID:g972931
C:Genetics:
A:Gene: IAA14
C:Superfamily: auxin-induced protein aux28
Query Match 55.0%; Score 44; DB 2; Length 164;
Best Local Similarity 61.5%; Pred. No. 2;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 2 DDDQWVYVQVWMP 14
DB 120 DGDWMLVGDVWMP 132
RESULT 2
H71407
auxin-induced protein - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
A:Variety: Columbia
C>Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 20-Jun-2000
C:Accession: H71407
R:Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp, R.; D
P.; Wedler, H.; Wedler, E.; Wambutt, R.; Weitzenegger, T.; Pohl, T.M.; Terry, N.; G
avanagh, T.; Hempel, S.; Kotter, P.; Entlan, K.D.; Rieger, M.; Schaeffer, M.; Funk, B
Nature 391, 485-488, 1998
A:Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.; Putdomen
ehoff, A.; Moores, T.; Jones, J.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.;
C.; Chalvatzis, N.
A>Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of Arabidopsis t
A:Reference number: A71400; MUID:98121113; PMID:9461215
A:Accession: H71407

G97360
Intracellular PHB depolymerase (AB017612) [Imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: G97360
R:Goodner, B.; Hinkle, G.; Gatlung, S.; Miller, N.; Blanchard, M.; Gurollo, B.; Goldman,
A.; Liu, F.; Mollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Matkeiz, B.;
Science 294, 2223-2326, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: G97360
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK85840.1; PID:G15154875; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_24
A:Map position: circular chromosome
C:Superfamily: Rickettsia prowazekii hypothetical protein RP681

Query Match 53.8%; Score 43; DB 2; Length 418;
Best Local Similarity 87.5%; Pred. No. 8.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 VVQVWVMP 14
11:111111
DB 242 VVQVWVMP 249

RESULT 8
AH2578
Intracellular PHB depolymerase [Imported] - Agrobacterium tumefaciens (strain C58, Dupont
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AH2578
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, L
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McClell
; Kemp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH2578
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-418 <KUR>
A:Cross-references: GB:AF006888; PIDN:AA141046.1; PID:G17738332; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: dephA
A:Map position: circular chromosome
C:Superfamily: Rickettsia prowazekii hypothetical protein RP681

Query Match 53.8%; Score 43; DB 2; Length 418;
Best Local Similarity 87.5%; Pred. No. 8.8;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 VVQVWVMP 14
11:111111
DB 242 VVQVWVMP 249

RESULT 9
T52251
5'-adenylylsulfate reductase (EC 1.8.99.-) [validated] - green alga (Enteromorpha intest
C:Species: Enteromorpha intestinalis (hollow green seaweed)
C:Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 20-Oct-2000
C:Accession: T52251
R:Guo, Y.; Schofield, O.M.; Leustek, T.
Plant Physiol. 123, 1087-1096, 2000
A:Title: Characterization of sulfate assimilation in marine algae focusing on the enzyme
A:Reference number: Z26003; MUID:20349720; PMID:10889256

A:Accession: T52251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-423 <GAO>
A:Cross-references: EMBL:AF069951; PIDN:AA26855.1
C:Function:
A:Description: catalyzes reduced glutathione-dependent reduction of adenylylsulfate t
C:Keywords: oxidoreductase

Query Match 53.8%; Score 43; DB 2; Length 423;
Best Local Similarity 58.3%; Pred. No. 8.9;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 DQDQVVVQVMP 13
1:111111
DB 330 DQDQVVVQVMP 341

RESULT 10
S51789
VLDL receptor precursor - chicken
N:Alternate names: very low density lipoprotein receptor; vitellogenin receptor
C:Species: Gallus gallus (chicken)
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 16-Jul-1999
C:Accession: S51789
R:Bujo, H.; Hermann, M.; Kaderli, M.O.; Jacobsen, L.; Sugawara, S.; Nimpf, J.; Yamano
EMBO J. 13, 5165-5175, 1994
A:Title: Chicken oocyte growth is mediated by an eight ligand binding repeat member o
A:Reference number: S51789; MUID:95045409; PMID:7957081
A:Accession: S51789
A:Molecule type: mRNA
A:Residues: 1-863 <BU>
A:Cross-references: EMBL:X80207; NID:G609265; PIDN:CA56505.1; PID:G609266
C:Comment: This receptor mediates uptake of very low density lipoprotein and vitellog
C:Comment: There is some evidence for the existence of a longer splice form contain
C:Keywords: duplication; fatty acid metabolism; glycoprotein; receptor; transmembrane
F:1-44/Domain: signal sequence #status predicted <SIG>
F:45-863/Product: VLDL receptor #status predicted <EXT>
F:45-787/Domain: extracellular #status predicted <EXT>
F:51-85/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:90-126/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:131-167/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:172-206/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:211-247/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:257-291/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:296-330/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:336-373/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:378-412/Domain: EGF homology <EG1>
F:418-452/Domain: EGF homology <EG2>
F:459-498/Domain: LDL receptor WTD-containing repeat homology <YW1>
F:499-544/Domain: LDL receptor WTD-containing repeat homology <YW2>
F:545-587/Domain: LDL receptor WTD-containing repeat homology <YW3>
F:588-631/Domain: LDL receptor WTD-containing repeat homology <YW4>
F:632-674/Domain: LDL receptor WTD-containing repeat homology <YW5>
F:675-717/Domain: LDL receptor WTD-containing repeat homology <YW6>
F:726-769/Domain: EGF homology <EG3>
F:788-809/Domain: transmembrane #status predicted <TM>
F:824-828/Region: coated-pit mediated internalization signal
F:159-773/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:378-389,385-398,400-412,418-428,424-437,439-452,726-735-734,736-769/Disulfide b

Query Match 53.8%; Score 43; DB 1; Length 863;
Best Local Similarity 35.7%; Pred. No. 20;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 1 YDQDQVVVQVMP 14
1:111111
DB 609 YDQDQVVVQVMP 622

RESULT 11
D86274

hypothetical protein F7A19.18 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
C:Accession: D86474
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzalla, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D86274
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-500 <STO>
A:Cross-references: GB:AE005172; NID:g5080784; PIDN:AA039294.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 52.5%; Score 42; DB 2; Length 500;
Best Local Similarity 50.0%; Pred. No. 16;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 2 DODQMVVQVWP 13
: ||||| ||||
Db 225 ESDQILIRQVWP 236

RESULT 12
T42584
tegument protein 40 - equine herpesvirus 4 (strain NS80567)
C:Species: equine herpesvirus 4
A:Variety: strain NS80567
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 21-Jul-2000
C:Accession: T42584
R:Teiford, E.A.; Watson, M.S.; Perry, J.; Cullinane, A.A.; Davison, A.J.
J. Gen. Virol. 79, 1197-1203, 1998
A>Title: The DNA sequence of equine herpesvirus-4
A:Reference number: Z22173; MUID:98264457; PMID:9603335
A:Accession: T42584
A:Molecule type: DNA
A:Residues: 1-529 <TEL>
A:Cross-references: EMBL:AF030027; NID:g2605950; PIDN:AAC59558.1; PID:g2605986
A:Experimental source: strain NS80567
C:Genetics:
A:Gene: 40
C:Superfamily: varicella-zoster virus gene 38 protein

Query Match 52.5%; Score 42; DB 2; Length 529;
Best Local Similarity 45.5%; Pred. No. 17;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 4 DQMVVQVWP 14
: ||||| ||||
Db 283 DEILIVEEPP 293

RESULT 13
C86468
probable auxin response factor, 53188-50111 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C86468
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C86468
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <STO>
A:Cross-references: GB:AE005172; NID:g10092376; PIDN:AA012783.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 52.5%; Score 42; DB 2; Length 600;
Best Local Similarity 53.8%; Pred. No. 20;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 DQDMVVVQVWP 14
: ||||| ||||
Db 566 DDDKMLVGDPP 578

RESULT 14
G86476
protein F1504.37 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86476
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Hulzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: G86476
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-767 <STO>
A:Cross-references: GB:AE005172; NID:g8778363; PIDN:AAF9371.1; GSPDB:GN00141
C:Genetics:
A:Gene: F1504.37
A:Map position: 1

Query Match 52.5%; Score 42; DB 2; Length 767;
Best Local Similarity 46.2%; Pred. No. 26;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 2 DQDMVVVQVWP 14
: ||||| ||||
Db 621 EEDKMLVGDPP 633

RESULT 15
WZBED4
gene 40 protein - equine herpesvirus 1 (strain Ab4p)
C:Species: equine herpesvirus 1
A:Note: host Equus caballus (domestic horse)
C>Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 16-Jul-1999
C:Accession: F36799
R:Teiford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
Submitted to GenBank, March 1992
A:Description: The DNA sequence of equine herpesvirus-1.
A:Reference number: A36805
A:Accession: F36799
A:Molecule type: DNA
A:Residues: 1-530 <TEL>
A:Cross-references: GB:M86664; NID:g330791; PIDN:AA02476.1; PID:g330833

R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.
 Virology 189, 304-316, 1992
 A:Title: The DNA sequence of equine herpesvirus-1.
 A:Reference number: A41831; MUID:92295566; PMID:1318606
 A:Contents: annotation; possible protein-coding frames
 A:Note: neither amino acid nor nucleotide sequence is given
 C:Genetics:
 A:Gene: 40
 C:Superfamily: varicella-zoster virus gene 38 protein

Query Match 51.28; Score 41; DB 1; Length 530;
 Best Local Similarity 36.48; Pred. No. 26;
 Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

Oy 4 DQVVVVQVPP 14
 I: : : : :
 Db 284 DEILIEPP 294

Search completed: March 27, 2003, 10:06:52
 Job time : 6.0122 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 2.30488 Seconds
(without alignments)
251.930 Million cell updates/sec

Title: US-09-675-208-3

Perfect score: 80

Sequence: 1 YDQDMVYGVWPW 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	44	55.0	229 1 AXIH_ARATH	P93830 arabidopsis
2	43	53.8	863 1 LQVR_CHICK	P98165 gallus gall
3	42	52.5	500 1 FQTB_ARATH	O9X178 arabidopsis
4	41	51.2	530 1 UL21_HSVB	P28972 equine herp
5	40	50.0	189 1 AXI6_ARATH	O38824 arabidopsis
6	40	50.0	653 1 ACSA_SYNY3	O54404 streptocyst
7	39	48.8	281 1 PHN2_ECOLI	P16688 escherichia
8	38	47.5	188 1 YECW_ECOLI	P52007 escherichia
9	38	47.5	404 1 WBS2_MOUSE	O54929 mus musculu
10	38	47.5	528 1 DYS2_MOUSE	O92630 mus musculu
11	38	47.5	574 1 ERG1_HUMAN	O14534 homo sapien
12	37	46.2	168 1 AXI2_ARATH	P46678 arabidopsis
13	37	46.2	174 1 AXI2_ARATH	P46678 arabidopsis
14	37	46.2	179 1 AXI6_PEA	P46680 pisum sativ
15	37	46.2	186 1 AXI4_ARATH	P33077 arabidopsis
16	37	46.2	188 1 AX2C_PHAU	O24541 phaseolus a
17	37	46.2	189 1 AXI4_PEA	O24541 phaseolus a
18	37	46.2	193 1 AX2D_PHAU	P46679 pisum sativ
19	37	46.2	194 1 AX2A_PHAU	P32293 phaseolus a
20	37	46.2	195 1 AX22_SOYB	P13088 glycine max
21	37	46.2	196 1 AX2B_PHAU	P33294 phaseolus a
22	37	46.2	203 1 AX2E_PHAU	O24541 phaseolus a
23	37	46.2	226 1 AX1G_ARATH	O24407 arabidopsis
24	37	46.2	243 1 AX17_ARATH	O38825 arabidopsis
25	37	46.2	243 1 AX17_ARATH	O38825 arabidopsis
26	37	46.2	251 1 SAST_ANAPL	P00633 anas platyr
27	37	46.2	261 1 AXIA_ARATH	O38826 arabidopsis
28	37	46.2	321 1 AXI8_ARATH	O38827 arabidopsis
29	37	46.2	338 1 AXI9_ARATH	O38827 arabidopsis
30	37	46.2	373 1 GSPL_XANCP	P34027 xanthomonas
31	37	46.2	428 1 GIN2_ORYSA	P14655 oryza sativ
32	37	46.2	475 1 DXR_MENPI	O9X60 mentha pipe
33	37	46.2	477 1 DXR_ARATH	O9X159 arabidopsis

34	37	46.2	532 1 GRB7_HUMAN	O14451 homo sapien
35	37	46.2	709 1 ANDR_RABIT	P49699 oryctolagus
36	37	46.2	873 1 LQVR_RABIT	P35953 oryctolagus
37	37	46.2	884 1 ANDR_EULFC	O97776 eulemur ful
38	37	46.2	895 1 ANDR_MACRA	O97952 macaca fasc
39	37	46.2	895 1 ANDR_PAPHA	O97960 papio hamad
40	37	46.2	899 1 ANDR_MOUSE	P15091 mus musculu
41	37	46.2	902 1 ANDR_RAT	P15207 rattus norv
42	37	46.2	907 1 ANDR_CANFA	O97775 canis famli
43	37	46.2	911 1 ANDR_PANTR	O97775 pan troglod
44	37	46.2	919 1 ANDR_HUMAN	P10275 homo sapien
45	37	46.2	1006 1 BGAL_LACDE	P20043 lactobacill

ALIGNMENTS

RESULT 1
AXIH_ARATH STANDARD: PRT: 229 AA.
AC P93830: 049162:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Auxin-responsive protein IAA17 (Indoleacetic acid-induced protein 17)
DE (Auxin response 3).
GN IAA17 OR AXR3 OR ATIG04250 OR P19P19.31.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=96004476; PubMed=9342315;
RA Kim J., Harter K., Theologis A.;
RT "Protein-protein interactions among the Aux/IAA proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:11786-11791(1997).
RN [2]
RP SEQUENCE FROM N.A., MUTANTS AXR3-1 AND AXR3-3, AND MUTAGENESIS OF
RP LEU-16; ASP-118 AND PRO-121.
RC STRAIN=cv. Columbia;
RX MEDLINE=96146427; PubMed=9478901;
RA Rouse D., Mackay P., Strinberg P., Estelle M., Leyser O.;
RT "Changes in auxin response from mutations in an AUX/IAA gene.";
RL Science 279:1371-1373(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altai H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
RA Dunn P., Etlgu P., Feldolym T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gail J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Tortum M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820(2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;

RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Toriumi M.J., Yu G., Brooks S.Y., Chao Q.,
 RA Chen H., Karlin-Neumann G., Kim C.J., Lam B., Miranda M., Nguyen M.,
 RA Palm C.J., Shinn P., Southwick A.M., Davis R.W., Ecker J.R.,
 RA Theologis A.,
 RT "Full Length cDNA of gene F19P19.31 (GI:4389514).";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT RIEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 RT SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 CC GROWTH (BY SIMILARITY).
 CC -1- SUBUNIT: Homo and heterodimers. Interacts with the auxin response
 CC factors ARF1 and IAA24.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- INDUCTION: BY AUXIN.
 CC -1- MISCELLANEOUS: Increased auxin response of mutants aux3-1 and
 CC aux3-3 may result form an increased stability of AUX3.
 CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL: U49073; AAB84354.1; -;
 CC DR EMBL: AF040631; AAC39439.1; -;
 CC DR EMBL: AF040632; AAC39440.1; -;
 CC DR EMBL: AC000104; AAB70451.2; -;
 CC DR EMBL: AF36916; AAG53997.1; -;
 CC DR EMBL: AY070094; AAL49831.1; -;
 CC DR TRANSFAC: T04501; -;
 CC DR InterPro: IPR003311; AUX_IAA.
 CC DR Pfam: PF02309; AUX_IAA.1.
 CC DR MultiGene Family: Nuclear protein; Translation regulation.
 CC FT MUTAGEN 88 P->L: IN AUX3-1; INCREASES AUXIN
 CC FT MUTAGEN 89 RESPONSE.
 CC FT MUTAGEN 89 V->G: IN AUX3-3; INCREASES AUXIN
 CC FT MUTAGEN 89 RESPONSE.
 CC FT MUTAGEN 16 L->F: SUPPRESSES THE AUX3-1 PHENOTYPE;
 CC FT MUTAGEN 16 WHEN ASSOCIATED WITH L-88.
 CC FT MUTAGEN 118 D->N: SUPPRESSES THE AUX3-1 PHENOTYPE;
 CC FT MUTAGEN 118 WHEN ASSOCIATED WITH L-88.
 CC FT MUTAGEN 121 P->S: SUPPRESSES THE AUX3-1 PHENOTYPE;
 CC FT MUTAGEN 121 WHEN ASSOCIATED WITH L-88.
 CC FT SEQUENCE 229 AA: 25288 MW: ECEA39207476581E CRC64;
 CC SQ
 CC
 CC Query Match 55.0%; Score 44; DB 1; Length 229;
 CC Best Local Similarity 61.5%; Pred. No. 1.1;
 CC Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 CC
 CC QY 2 DDDQVVVVVYVWPP 14
 CC Db 185 DGDMMLVGDVWPP 197
 CC
 CC RESULT 2
 CC LDVR_CHICK
 CC ID LDVR_CHICK STANDARD: PRT; 863 AA.
 CC AC P98165;
 CC DT 01-OCT-1996 (Rel. 34, Created)
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Very low-density lipoprotein precursor (VLDL receptor)
 CC (Vitellogenin receptor) (Vtg receptor).

GN VLDLR OR VTGR.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OC NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=white leghorn; TISSUE=ovary;
 RX MEDLINE=95045409; PubMed=7957081;
 RA Bujo H., Hermann M., Kaderli M.O., Jacobsen L., Sugawara S.,
 RA Nimpt J., Yamamoto T., Schneider W.J.;
 RT "Chicken oocyte growth is mediated by an eight ligand binding repeat
 RT member of the LDL receptor family.";
 RL EMBO J. 13:5165-5175(1994).
 RN [2]
 RN SEQUENCE OF 510-518; 546-554 AND 819-827.
 RP STRAIN=white leghorn; TISSUE=follicle membrane;
 RX MEDLINE=92011638; PubMed=1655760;
 RA Barber D.L., Sanders E.J., Abersold R., Schneider W.J.;
 RT "The receptor for yolk lipoprotein deposition in the chicken oocyte.";
 RL J. Biol. Chem. 266:18761-18770(1991).
 CC -1- FUNCTION: Binds VLDL and transports it into cells by endocytosis.
 CC In order to be internalized, the receptor-ligand complexes must
 CC first cluster into clathrin-coated pits. Binding to Reelin induces
 CC tyrosine phosphorylation (By similarity).
 CC -1- SUBUNIT: Binds to the extracellular matrix protein Reelin (by
 CC similarity).
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- TISSUE SPECIFICITY: ABUNDANT IN OOCYTES; MUCH LESS IN HEART AND
 CC SKELETAL MUSCLE.
 CC -1- SIMILARITY: CONTAINS 6 LDL-RECEPTOR CLASS B DOMAINS.
 CC -1- SIMILARITY: CONTAINS 3 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC DR EMBL: X80207; CAA56505.1; -;
 CC DR HSP: P01130; IADJ.
 CC DR InterPro: IPR000152; Asx_hydroxyl.
 CC DR InterPro: IPR000561; EGF-like.
 CC DR InterPro: IPR001861; EGF_CA.
 CC DR InterPro: IPR002172; LDL_recept_A.
 CC DR InterPro: IPR000033; LDL_recept_rep.
 CC DR Pfam: PF00057; ldl_recept_a; 8
 CC DR Pfam: PF00058; ldl_recept_b; 5.
 CC DR PRINTS: PRO0261; LDLRECEPTOR.
 CC DR SMART: SM00179; EGF_CA; 2.
 CC DR SMART: SM00001; EGF-like; 1.
 CC DR SMART: SM00192; LDLra; 8.
 CC DR SMART: SM00135; LY; 5.
 CC DR PROSITE: PS00010; ASX_HYDROXYL; 2.
 CC DR PROSITE: PS00022; EGF_1; FALSE-NEG.
 CC DR PROSITE: PS01186; EGF_2; 3.
 CC DR PROSITE: PS01187; EGF_CA; 2.
 CC DR PROSITE: PS01209; LDLRA_1; 8.
 CC DR PROSITE: PS00068; LDLRA_2; 8.
 CC DR Glycoprotein; VLDL; Cholesterol metabolism; Lipid transport;
 CC Endocytosis; Coated pits; Transmembrane; Receptor; Signal;
 CC KW EGF-like domain; Repeat.
 CC KW SIGNAL 1 43
 CC FT CHAIN 44 863
 CC FT DOMAIN 44 785
 CC FT TRANSMEM 786 809
 CC FT DOMAIN 810 863
 CC FT DOMAIN 49 87
 CC
 CC POTENTIAL.
 CC VERY LOW-DENSITY LIPOPROTEIN RECEPTOR.
 CC EXTRACELLULAR (POTENTIAL).
 CC POTENTIAL.
 CC CYTOPLASMIC (POTENTIAL).
 CC LDL-RECEPTOR CLASS A 1.

```

FT DOMAIN 88 128 LDL-RECEPTOR CLASS A 2.
FT DOMAIN 129 169 LDL-RECEPTOR CLASS A 3.
FT DOMAIN 170 208 LDL-RECEPTOR CLASS A 4.
FT DOMAIN 209 249 LDL-RECEPTOR CLASS A 5.
FT DOMAIN 255 293 LDL-RECEPTOR CLASS A 6.
FT DOMAIN 294 332 LDL-RECEPTOR CLASS A 7.
FT DOMAIN 334 373 LDL-RECEPTOR CLASS A 8.
FT DOMAIN 374 413 EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 414 453 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT REPEAT 457 498 LDL-RECEPTOR CLASS B 1.
FT REPEAT 499 544 LDL-RECEPTOR CLASS B 2.
FT REPEAT 545 587 LDL-RECEPTOR CLASS B 3.
FT REPEAT 588 631 LDL-RECEPTOR CLASS B 4.
FT REPEAT 632 674 LDL-RECEPTOR CLASS B 5.
FT REPEAT 675 716 LDL-RECEPTOR CLASS B 6.
FT DOMAIN 722 770 EGF-LIKE 3.
FT SITE 822 827 ENDOTOXIN SIGNAL (POTENTIAL).
FT DISULFID 51 63 BY SIMILARITY.
FT DISULFID 58 76 BY SIMILARITY.
FT DISULFID 70 85 BY SIMILARITY.
FT DISULFID 90 102 BY SIMILARITY.
FT DISULFID 97 115 BY SIMILARITY.
FT DISULFID 109 126 BY SIMILARITY.
FT DISULFID 131 145 BY SIMILARITY.
FT DISULFID 138 158 BY SIMILARITY.
FT DISULFID 152 167 BY SIMILARITY.
FT DISULFID 172 184 BY SIMILARITY.
FT DISULFID 179 197 BY SIMILARITY.
FT DISULFID 191 206 BY SIMILARITY.
FT DISULFID 211 223 BY SIMILARITY.
FT DISULFID 223 236 BY SIMILARITY.
FT DISULFID 230 247 BY SIMILARITY.
FT DISULFID 257 269 BY SIMILARITY.
FT DISULFID 264 282 BY SIMILARITY.
FT DISULFID 276 291 BY SIMILARITY.
FT DISULFID 296 308 BY SIMILARITY.
FT DISULFID 303 321 BY SIMILARITY.
FT DISULFID 315 330 BY SIMILARITY.
FT DISULFID 336 349 BY SIMILARITY.
FT DISULFID 344 362 BY SIMILARITY.
FT DISULFID 356 373 BY SIMILARITY.
FT DISULFID 378 389 BY SIMILARITY.
FT DISULFID 385 398 BY SIMILARITY.
FT DISULFID 400 412 BY SIMILARITY.
FT DISULFID 418 428 BY SIMILARITY.
FT DISULFID 424 437 BY SIMILARITY.
FT DISULFID 439 452 BY SIMILARITY.
FT DISULFID 439 452 BY SIMILARITY.
FT DISULFID 726 739 BY SIMILARITY.
FT DISULFID 735 754 BY SIMILARITY.
FT CARBOHYD 756 769 BY SIMILARITY.
FT CARBOHYD 169 169 N-LINKED (GLUCAN. . .) (POTENTIAL).
FT CARBOHYD 773 773 N-LINKED (GLUCAN. . .) (POTENTIAL).
SO SEQUENCE 863 AA: 94904 MW: 0672AB74BF9A2245 CRC64;

```

```

Query Match 53.8%; Score 43; DB 1; Length 863;
Best Local Similarity 35.7%; Pred. No. 7.1;
Matches 5; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```

```

Oy 1 YDQDMVVVQVWP 14
Db 609 FDRQDLVTTEIOWP 622

```

RESULT 3

```

FUT8_ARATH STANDARD: PRT: 500 AA.

```

```

AC 09X178;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE Probable fucosyltransferase 8 (EC 2.4.1.-) (AtFUT8).
GN FUT8 OR AT1G14100 OR F7A19.18.
OS Arabidopsis thaliana (Mouse-ear cress).

```

```

CC Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
CC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
CC NCBI_TaxID=3702;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=cv. Columbia;
CC MEDLINE=21016719; PubMed=11130712;
CC White O., Alonso J., Altabi H., Araujo R., Bowman C.L., Brooks S.Y.,
CC Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
CC Chung M.K., Conn L., Conway A.B., Conway T.H., Dewar K.,
CC Dunn P., Elgu P., Feldblum T.V., Feng J.-D., Fong B., Fujil C.Y.,
CC Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
CC Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
CC Kim C.J., Koo H.L., Kremenetskaia I., Kuriz D.B., Kwan A., Lam B.,
CC Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
CC Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
CC Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
CC Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
CC Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
CC Sun H., Tallon L.J., Rambunga G., Tortum M.J., Town C.D.,
CC Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
CC Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
CC "Sequence and analysis of chromosome 1 of the plant Arabidopsis
CC thaliana."
CC Nature 408:816-820(2000).
CC [2]
CC IDENTIFICATION AS PUTATIVE FUCOSYLTRANSFERASE, AND TISSUE SPECIFICITY.
CC MEDLINE=21608393; PubMed=11743104;
CC Sarría R., Wagner T.A., O'Neill M.A., Falk A., Milkersen C.G.,
CC Keegstra K., Ralshel N.V.;
CC "Characterization of a family of Arabidopsis genes related to
CC xyloglucan fucosyltransferase1."
CC plant Physiol. 127:1595-1606(2001).
CC - FUNCTION: May be involved in cell wall biosynthesis. May
CC act as a fucosyltransferase.
CC - PATHWAY: Glycosylation.
CC - SUBCELLULAR LOCATION: Golgi (Potential).
CC - TISSUE SPECIFICITY: Expressed in leaves and stems.
CC - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 37.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC007576; AAC39294.1; -
CC InterPro: IPR004938; XG_FTHase.
CC Pfam: PF03254; XG_FTHase.1
CC Transferrase; Glycosyltransferase; Golgi stack; Cell wall.
CC KW SEQUENCE 500 AA: 57538 MW: 1C5BB6152BFAE690 CRC64;

```

```

Query Match 52.5%; Score 42; DB 1; Length 500;
Best Local Similarity 50.0%; Pred. No. 6;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

```

```

Oy 2 DQDMVVVQVWP 13
Db 225 ESDQILIRQVWP 236

```

RESULT 4

```

UL21_HSVB STANDARD: PRT: 530 AA.

```

```

AC P28972;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DE Gene 40 protein.
GN Gene 40 protein.
OS

```

```

GN 40.
OS Equine herpesvirus type 1 (strain Ab4p) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1."
RL Virology 189:304-316(1992).
CC -1- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL21,
CC EHV-1 40, EHV-4 UL21, AND VZV 38.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M86664; AAB02476.1; -
CC PIR: F36799; WZBEDA.
DR InterPro: IPR004936; Herpes_UL21.
DR Pfam: PF03252; UL21; 1.
SQ SEQUENCE 530 AA; 57915 MW; 2E43565AB297CD18 CRC64;

Query Match
Best Local Similarity 36.4%; Pred. No. 9.7; Length 530;
Matches 4; Conservative 6; Mismatches 1; Indels 0; Gaps 0;

OY 4 DQMVVQVQVMP 14
ID 284 DEILIEEPPMP 294

RESULT 5
AA16_ARATH STANDARD; PRT; 189 AA.
AC 038824; OGC539;
ID 16-OCT-2001 (Rel. 40, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DI 15-JUN-2002 (Rel. 41, Last annotation update)
DE Auxin-responsive protein IAA6 (Indoleacetic acid-induced protein 6).
GN IAA6 OR AIG52830 OR FI4G24.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RA MEDLINE=95387393; PubMed=7658471;
RA Abel S., Nguyen M.D., Theologis A.;
RT "The PS-1AA4/5-like family of early auxin-inducible mRNAs in
RT Arabidopsis thaliana."
RL J. Mol. Biol. 251:533-549(1995).

RP SEQUENCE FROM N.A.
RX STRAIN=cv. Columbia;
RA MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
RA Dunn P., Elgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huitzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Marshall A.,

```

```

RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.,
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana."
RL Nature 408:816-820(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA Goldsmith A.D., Lee J.M., Toriumi M., Yu G., Brooks S., Chao Q.,
RA Chen H., Karlin-Neumann G., Kim C., Lam B., Miranda M., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene FI4G24.10 (GI:6862944)."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
CC GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U18408; AAC49047.1; -
CC DR EMBL: AC019018; AAG52268.1; -
CC DR EMBL: AF336915; AAG53996.1; -
CC DR InterPro: IPR003311; AUX_IAA.
CC DR Pfam: PF02309; AUX_IAA; 1.
CC KM Multigene family; Nuclear protein; translation regulation.
CC FT CONFIRM 34 34 K -> N (IN REF. 1).
SQ SEQUENCE 189 AA; 21031 MW; 5EB9130584A75465 CRC64;

Query Match
Best Local Similarity 50.0%; Score 40; DB 1; Length 189;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 2 DQDQVVQVQVMP 13
ID 152 DRDWMVLVGDPMP 163

RESULT 6
ACSA_SYNY3 STANDARD; PRT; 653 AA.
AC 055404;
DI 30-MAY-2000 (Rel. 39, Created)
DI 30-MAY-2000 (Rel. 39, Last sequence update)
DI 16-OCT-2001 (Rel. 40, Last annotation update)
DE Acetyl-coenzyme A synthetase (EC 6.2.1.1) (acetate--CoA ligase) (Acyl-
DE activating enzyme).
GN ACS OR SLI0542.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96127529; PubMed=8590279;
RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
RA Sugita M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
RT region from map positions 648 to 928 of the genome."
RL DNA Res. 2:153-166(1995).

```



```

CC -1- CATALYTIC ACTIVITY: ATP + acetate + CoA -> AMP + diphosphate +
CC acetyl-CoA.
CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
CC FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64003; BAAL0498.1; -.
DR HSSP; P08659; 1LCI.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding.1.
DR PRINTS; PR00154; AMPBINDING.
DR PROSITE; PS00455; AMP_BINDING.1.
KM L19aag; Complete proteome.
SQ SEQUENCE 653 AA; 73052 MW; 5EBCAD430EEFFA03 CRC64;

Query Match 50.0%; Score 40; DB 1; Length 653;
Best Local Similarity 46.2%; Pred. No. 18;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

OY 2 DQDMVVQVWP 14
   11 111111
Db 464 DQGFVITKQPM 476

RESULT 7
ID PHNJ_ECOLI STANDARD: PRT: 281 AA.
AC P16688;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PhnJ protein.
GN PHNJ OR B4098.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91193228; PubMed=1840580;
RA Makino K., Kim S.K., Shingawa H., Amemura M., Nakata A.;
RT "Molecular analysis of the cryptic and functional phn operons for
RT phosphate use in Escherichia coli K-12.";
RL Bacteriol. 173:2665-2672(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=95334362; PubMed=7610040;
RA Burland V.D., Plunkett G. III, Sofia H.J., Daniels D.L.,
RA Blattner F.R.;
RT "Analysis of the Escherichia coli genome VI: DNA sequence of the
RT region from 92.8 through 100 minutes.";
RL Nucleic Acids Res. 23:2105-2119(1995).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=B;
RX MEDLINE=90170953; PubMed=2155230;
RA Chen C.-M., Ye Q.-Z., Zhu Z., Warner B.L., Walsh C.T.;
RT "Molecular biology of carbon-phosphorus bond cleavage. Cloning and
RT sequencing of the phn (psd) genes involved in alkyldiphosphate
RT uptake and C-P lyase activity in Escherichia coli B.";
RL J. Biol. Chem. 265:4461-4471(1990).
CC -1- FUNCTION: BELONGS TO AN OPERON INVOLVED IN ALKYLPHOSPHONATE
CC UPTAKE AND C-P LYASE. EXACT FUNCTION NOT KNOWN.
CC -1- MISCELLANEOUS: THE SEQUENCE SHOWN IS THAT OF STRAIN K12.

```

```

CC -1- SIMILARITY: BELONGS TO THE PHNJ FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D90227; BAAL4270.1; -.
DR EMBL; U14003; AAA96997.1; -.
DR EMBL; AE000482; AAC77059.1; -.
DR EMBL; J05260; AAA24348.1; -.
DR PIR; B35719; B35719.
DR EcoGene; EG10719; phnJ.
KM Alkylphosphonate uptake; Complete proteome.
FT VARIANT 103 103 V -> L (IN STRAIN B).
SQ SEQUENCE 281 AA; 31845 MW; 241F6AF140995468 CRC64;

Query Match 48.8%; Score 39; DB 1; Length 281;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

OY 3 QDMVVQVWP 14
   11111111
Db 117 EDQIIFQVPI 128

RESULT 8
ID YECC_ECOLI STANDARD: PRT: 188 AA.
AC P52007; P52008; P76295;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein yecM.
GN YECC OR B1875.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-VIDES J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T.,
RA Isono K., Kasai H., Kimura S., Kitakawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshino H.,
RA Saito N., Sampei G., Seki Y., Sivasubramanian S., Tagami H.,
RA Takeda J., Takemoto K., Wada C., Yamamoto Y., Horikuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
RN [3]
RP SEQUENCE OF 102-188 FROM N.A.
RC STRAIN=K12 / MC4100;
RX Gupta S.D., Wu H.C.;
RT Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.
RN [4]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=99420866; PubMed=10493123;

```

```

RA Fountoulakis M., Takacs M.-F., Berndt P., Langen H., Takacs B.;
RT "Enrichment of low abundance proteins of Escherichia coli by
RT hydroxapatite chromatography."
RL Electrophoresis 20:2181-2195(1999).
CC -1- SIMILARITY: STRONG, TO H.INFLOENZAE H11582/H11581.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 128 THAT PRODUCES TWO SEPARATE ORFS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000281; AAC74945.1; ALT_INIT.
DR EMBL; D90829; BAA15685.1; -.
DR EMBL; D90830; BAA15692.1; -.
DR EMBL; L38618; AAA89200.1; ALT_FRAME.
DR EMBL; L38618; AAA89201.1; ALT_FRAME.
DR EcoGene; EG13405; yecm.
KW Complete proteome.
FT FT 102 G -> R (IN REF. 3).
FT CONFLICT 120 G -> V (IN REF. 3).
FT CONFLICT 167 G -> A (IN REF. 3).
SQ SEQUENCE 188 AA; 21205 MW; 700BBE042FFAA8A5 CRC64;

Query Match 47.5%; Score 38; DB 1; Length 188;
Best Local Similarity 50.0%; Pred. No. 11;
Matches 5; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Oy 5 QMVVVQVWP 14
Db 92 QFSIVLPWP 101
I:::|||

RESULT 9
WBS2_MOUSE STANDARD; PRT; 404 AA.
ID WBS2_MOUSE
AC 054929; Q9RJM8;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE WD repeat and SOCS box containing protein 2 (WSB-2) (SOCS box-
DE containing WD protein SWIP-2).
DE WSB2 OR SWIP2.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RF STRAIN=C57BL/6; TISSUE=Spleen;
RC MEDLINE=98081836; PubMed=9419338;
RA Hilton D.J., Richardson R.T., Alexander W.S., Viney E.M.,
RA Milson T.A., Sprig N.S., Starr R., Nicholson S.E., Metcalf D.,
RA Nicola N.A.;
RT "Twenty proteins containing a C-terminal SOCS box form five structural
RT classes."
RL Proc. Natl. Acad. Sci. U.S.A. 95:114-119(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99284520; PubMed=10354473;
RA Vasilioukas D., Hancock S., Stern C.D.;
RT "SWIP-1: novel SOCS box containing WD-protein regulated by signalling
RT centres and by Shh during development."
RL Mech. Dev. 82:79-94(1999).
CC -1- SIMILARITY: CONTAINS 5 WD REPEATS (TRP-ASP DOMAINS).
CC -1- SIMILARITY: CONTAINS 1 SOCS BOX DOMAIN
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF033188: AAB96649.1; -.
DR EMBL: AF072881: AAD28809.1; -.
DR InterPro: IPR001496: SOCS.
DR InterPro: IPR001680: WD40.
DR Pfam: PF00400: WD40. 6.
DR PRINTS: PR00320: GPROTEINBRPT.
DR ProDom: PD000018: WD40. 2.
DR SMART: SM00253: SOCS. 1.
DR SMART: SM00320: WD40. 6.
DR PROSITE: PS00678: WD_REPEATS_1; 2.
DR PROSITE: PS50082: WD_REPEATS_2; 5.
DR PROSITE: PS50294: WD_REPEATS_REGION; 1.
DR PROSITE: PS50225: SOCS. 1.
DR Repeat: WD repeat.
DR Repeat: WD 1.
FT REPEAT 151 191 WD 2.
FT REPEAT 195 234 WD 3.
FT REPEAT 237 276 WD 4.
FT REPEAT 291 330 WD 5.
FT DOMAIN 356 404 SOCS BOX.
FT CONFLICT 272 272 A -> E (IN REF. 2).
SQ SEQUENCE 404 AA; 45203 MW; 78F666FAF9AC6659A CRC64;

Query Match 47.5%; Score 38; DB 1; Length 404;
Best Local Similarity 50.0%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

OY 1 YDQDMVVVQVWP 14
Db 43 WSQCHGVKLVWP 56

RESULT 10
DYN2_HUMAN
ID DYN2_HUMAN STANDARD; PRT; 528 AA.
AC Q92630;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Dual-specificity tyrosine-phosphorylation regulated kinase 2
DE (EC 2.7.1.-).
GN DYRK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=98421512; PubMed=9748265;
RA Becker W., Weber Y., Wetzel K., Eimbter K., Tejedor F.J.,
RA Joost H.-G.;
RA "Sequence characteristics, subcellular localization, and substrate
RT specificity of DYRK-related kinases, a novel family of dual
RT specificity protein kinases.";
RL J. Biol. Chem. 273:25893-25902(1998).
RN [2]
RP SEQUENCE OF 320-528 FROM N.A.
RC TISSUE=Placenta;
RC Becker W., Joost H.-G.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: IN VITRO, CAN PHOSPHORYLATE HISTONES H3 AND H2B ON SER
CC AND THR RESIDUES. MAY BE INVOLVED IN THE REGULATION OF CELLULAR
CC GROWTH AND/OR DEVELOPMENT.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- PM: AUTHOPHOSPHORYLATED ON TYR RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

```

```

CC -----
CC MMB/DYRK SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y13493; CA73885.1; -.
CC EMBL: Y09216; CA70418.1; -.
CC HSSP: Q00534; 1B18.
CC Genew: HGNC:3093; DYRK2.
CC MIM: 603496; -.
CC InterPro: IPR000719; Euk_pkinase.
CC InterPro: IPR002290; Ser_thr_pkinase.
CC Pfam: PF00069; pkinase; 1.
CC ProDom: PD000001; Euk_pkinase; 1.
CC SMART: SM00220; S_TKC; 1.
CC PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
CC PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
CC PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
CC Transferrase; Serine/threonine-protein kinase; Tyrosine-protein kinase;
CC ATP-binding; Phosphorylation.
CC KW DOMAIN 149 462
CC NP_BIND 155 163 ATP (BY SIMILARITY).
CC BINDING 178 178 ATP (BY SIMILARITY).
CC ACT_SITE 275 275 BY SIMILARITY.
CC SO SEQUENCE 528 AA; 59714 MW; AF2C682ED9522D7 CRC64;

Query Match 47.5%; Score 38; DB 1; Length 528;
Best Local Similarity 66.7%; Pred. No. 33;
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 YDQDMVYVQVP 12
Db 130 YDDDCSYVQVP 141

RESULT 11
ERGL_HUMAN STANDARD; PRT; 574 AA.
AC Q14534; Q9UEK6;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Squalene monooxygenase (EC 1.14.99.7) (Squalene epoxidase) (SE).
GN SQUE OR ERGL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE:97432831; PubMed-9286711;
RA Nagai M., Sakakibara J., Wakui K., Fukushima Y., Igarashi S.,
RA Tani S., Arakawa M., Ono T.;
RT "Localization of the squalene epoxidase gene (SOLE) to human
RT chromosome region Bq24.1."
RT Genomics 44:141-143(1997).
RN [2]
RP SEQUENCE OF 187-535 FROM N.A.
RC TISSUE=Liver;
RX MEDLINE:96215195; PubMed-8626488;
RA Nakamura Y., Sakakibara J., Izumi T., Shibata A., Ono T.;
RT "Transcriptional regulation of squalene epoxidase by sterols and
RT inhibitors in HeLa cells."
RT J. Biol. Chem. 271:8053-8056(1996).
CC -1- FUNCTION: CATALYZES THE FIRST OXYGENATION STEP IN STEROL
CC BIOSYNTHESIS AND IS SUGGESTED TO BE ONE OF THE RATE-LIMITING
CC ENZYMES IN THIS PATHWAY.
CC -1- CATALYTIC ACTIVITY: Squalene + AH(2) + O(2) = (S)-squalene-2,3-
```

```

CC -----
CC epoxide + A + H(2)O.
CC -1- COFACTOR: FAD.
CC -1- SUBUNIT: MAY FORM A COMPLEX WITH SQUALENE SYNTHASE.
CC -1- SUBCELLULAR LOCATION: Mitochondrion.
CC -1- SIMILARITY: BELONGS TO THE SQUALENE MONOOXYGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D78130; BAA22372.1; -.
CC EMBL: D78129; BAA11209.1; -.
CC Genew: HGNC:11279; SOLE.
CC MIM: 602019; -.
CC InterPro: IPR000733; Flav_monooxygenase.
CC InterPro: IPR000205; NAD_binding.
CC InterPro: IPR003042; Rng_monooxygenase.
CC Pfam: PF01360; Monooxygenase; 1.
CC PRINTS: PR00420; RINGMOXGNASE.
CC KW Oxidoreductase; Flavoprotein; FAD; Transmembrane.
CC TRANSMEM 20 40 POTENTIAL.
CC TRANSMEM 61 81 POTENTIAL.
CC TRANSMEM 123 143 POTENTIAL.
CC TRANSMEM 546 566 POTENTIAL.
CC NP_BIND 126 153 FAD (ADP PART) (POTENTIAL).
CC CONFLICT 247 247 E -> G (IN REF. 2).
CC CONFLICT 336 336 R -> Q (IN REF. 2).
CC CONFLICT 389 389 L -> P (IN REF. 2).
CC CONFLICT 451 451 N -> K (IN REF. 2).
CC CONFLICT 518 518 A -> V (IN REF. 2).
CC SO SEQUENCE 574 AA; 63939 MW; DDAFP95308864FD09 CRC64;

Query Match 47.5%; Score 38; DB 1; Length 574;
Best Local Similarity 42.9%; Pred. No. 36;
Matches 6; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 1 YDQDMVYVQVP 14
Db 198 HDQSKSEVQIDYP 211

RESULT 12
AXIL_ARATH STANDARD; PRT; 168 AA.
AC P49677; O23312;
DT 01-FEB-1996 (Rel. 33, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Auxin-responsive protein IAA1 (Indoleacetic acid-induced protein 1).
GN IAA1 OR A74G14560 OR DIJ320W.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_Taxid:3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE:94105161; PubMed-8278386;
RA Abel S., Oeller P.W., Theologis A.;
RT "Early auxin-induced genes encode short-lived nuclear proteins."
RT Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE:20083488; PubMed-10617198;
RA Mayer K.F.X., Schueller C., Wambolt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Ertlan K.-D., Terry N.,
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
```

RA Weichselgartner M., de Simone V., Obermaier B., Macho R., Mueller M.,
 RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidthein T.,
 RA Reichert B., Portetle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hebeisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grynoprez B., Chuang Y.-J., Vandebussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Deloor E.,
 RA Weltenegeger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirke W.,
 RA Moellman P., Klein lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Bernieris S., Hempel S., Feldpausch M., Lambirth S., Van den Daele H.,
 RA De Keyser A., Buysbaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail A., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McIay K., Mayes R.,
 RA Petlett A., Rajandream M.A., Lyne M., Benes V., Reckmann S.,
 RA Borikova D., Bloecker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Heitzl A.,
 RA Neumann S., Argitrou A., Vitale D., Liguori R., Pitravadi E.,
 RA Massenet O., Quigley F., Clabaud G., Muendlein A., Feilber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stoecker H.,
 RA Zaccaria P., Bevan M., Willson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedha N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoeckling T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antonou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
 RA Granat S., Shoudy N., Hasegawa A., Hamed A., Lochl M., Johnson A.,
 RA Chen E., Maria M., Martienssen R., McCombie W.R.,
 RT "Sequence and analysis of chromosome 4 of the plant *Arabidopsis thaliana*."
 RT Nature 402:769-777(1999).
 RL
 CC -!- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 CC GROWTH.
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 CC GROWTH.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- INDUCTION: BY AUXIN.
 CC -!- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L15448; AAA16569.1; -
 CC DR EMBL: 297336; CAB10235.1; -
 CC DR EMBL: AL161539; CAB78498.1; -
 CC DR TRANSFAC: T04508; -
 CC DR InterPro: IPR003311; AUX_IAA.
 CC DR Pfam: PF02309; AUX_IAA.1.
 CC MultiGene family: Nuclear protein; Translation regulation.
 FT CONFLICT 52 52 A -> T (IN REF. 1).
 SO SEQUENCE 168 AA; 19031 MW; 17CA9B3BBE7962EE CRC64;

QY 2 DDDQWVYVPM 13

Query Match 46.2%; Score 37; DB 1; Length 168;
 Best Local Similarity 58.3%; Prec. No. 15;
 Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

DB 135 DGDMLVGVPM 146
 RESULT 13
 AX12_ARATH STANDARD; PRT; 174 AA.
 AC P49678; O22596;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Auxin-responsive protein IAA2 (indoleacetic acid-induced protein 2).
 GN IAA2 OR AT3G23030 OR MXC7.6.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=94105161; PubMed=8278386;
 RA Abel S., Oeller P.W., Theologis A.;
 RT "Early auxin-induced genes encode short-lived nuclear proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Wassilewskija;
 RA Malik M.K., Zimmerman J.L.;
 RT "Molecular characterization of a genomic clone encoding the auxin-
 RT inducible IAA2 gene from *Arabidopsis thaliana*."
 RL (in) Plant Gene Register PGR97-178.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=20277480; PubMed=10819329;
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Tabata S.;
 RT "Structural analysis of *Arabidopsis thaliana* chromosome 3. I. Sequence
 RT features of the regions of 4,504,864 bp covered by sixty P1 and YAC
 RT clones.";
 RL DNA Res. 7:131-135(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Egu P., Lee J.M.,
 RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Karlin-Neumann G.,
 RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shinn P.,
 RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene MXC7.6 (GI:9294192)."
 CC Submitted (Dec-2000) to the EMBL/Genbank/DBJ databases.
 CC -!- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
 CC GROWTH.
 CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
 CC GROWTH.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- INDUCTION: BY AUXIN.
 CC -!- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L15449; AAA16570.1; -
 CC DR EMBL: AF027157; AAB97164.1; -
 CC DR EMBL: AB026655; BAB02094.1; -
 CC DR EMBL: AF332392; AAG48756.1; -
 CC DR TRANSFAC: T04509; -
 CC DR InterPro: IPR003311; AUX_IAA.
 CC DR Pfam: PF02309; AUX_IAA.1.
 CC MultiGene family: Nuclear protein; Translation regulation.

```

FT  CONFLICT 38 38 S -> I (IN REF. 2).
SQ  SEQUENCE 174 AA: 19910 MW: DECA4F59F0106179 CRC64;

Query Match
Best Local Similarity 58.3%; Score 37; DB 1; Length 174;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DODOMVVVQVPW 13
    1 1 1 1 1 1 1
DB 138 DGDMLVGVDPW 149

RESULT 14
AX16_PEA STANDARD: PRT; 179 AA.
AC P49680:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Auxin-induced protein IAA6.
GN IAA6.
OS Pisum sativum (Garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
OX NCBI_TaxID=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Alaska;
RX MEDLINE=94016594; PubMed=8411182;
RA Oeller P.W., Keller J.A., Parks J.E., Silbert J.E., Theologis A.;
RT "Structural characterization of the early indoleacetic acid-inducible
RT genes, PS-IAA4/5 and PS-IAA6, of pea (Pisum sativum L.).";
RL J. Mol. Biol. 233:789-798(1993).
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CC GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: X68218; CAA48300.1; -.
CC DR EMBL: X68217; CAA48299.1; -.
CC DR InterPro: IPR003311; AUX_IAA.
CC Pfam: PF02309; AUX_IAA; 1.
CC Multigene family: Nuclear protein; Translation regulation.
CC SO SEQUENCE 179 AA: 20330 MW: 0D212F384D100D7E CRC64;

Query Match
Best Local Similarity 58.3%; Score 37; DB 1; Length 179;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DODOMVVVQVPW 13
    1 1 1 1 1 1 1
DB 135 DGDMLVGVDPW 146

RESULT 15
AX14_ARATH STANDARD: PRT; 186 AA.
AC P33077:
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Auxin-responsive protein IAA4 (indoleacetic acid-induced protein 4)

```

```

DE (Auxin-induced protein AUX2-11).
GN IAA4 OR AUX2-11 OR AT5G43700 OR MOD19.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=91338700; PubMed=2102379;
RA Conner T.W., Goekjian V.H., Lafayette P.R., Key J.L.;
RT "Structure and expression of two auxin-inducible genes from
RT Arabidopsis.";
RL Plant Mol. Biol. 15:623-632(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=94105161; PubMed=8278386;
RA Abel S., Oeller P.W., Theologis A.;
RT "Early auxin-induced genes encode short-lived nuclear proteins.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:326-330(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Eryu P., Lee J.M.,
RA Toriumi M., Yu G., Brooks S., Chao Q., Chen H., Katlin-Neumann G.,
RA Kim C., Lam B., Miranda M., Nguyen M., Palm C.J., Shin P.,
RA Southwick A., Davis R.W., Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MOD19.3 (GI:10177938).";
RT Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: COULD ACT AS REGULATOR OF GENES RESPONSIBLE FOR
CC MEDIATING THE VARIOUS AUXIN-INDUCED EVENTS RESPONSIBLE FOR CELL
CC GROWTH.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY AUXIN.
CC -1- SIMILARITY: BELONGS TO THE AUX/IAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC DR EMBL: X53435; CAA37526.1; -.
CC DR EMBL: L15450; AAA16571.1; -.
CC DR EMBL: AB026651; BAB11297.1; -.
CC DR EMBL: AF332394; AAC48758.1; -.
CC DR PIR: S12243; S12243.
CC TRANSFAC: T04511; -.
CC InterPro: IPR003311; AUX_IAA.
CC Pfam: PF02309; AUX_IAA; 1.
CC Multigene family: Nuclear protein; Translation regulation.
CC SO SEQUENCE 186 AA: 20975 MW: DCF66CAC4D28CD3B CRC64;

Query Match
Best Local Similarity 58.3%; Score 37; DB 1; Length 186;
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

OY 2 DODOMVVVQVPW 13
    1 1 1 1 1 1 1
DB 149 DGDMLVGVDPW 160

```

Fri Apr 4 12:31:26 2003

us-09-675-208-3.rsp

Page 10

Search completed: March 27, 2003, 10:04:18
Job time : 3.30488 secs

```

DE Rafinose synthase protein.
GN Mp012.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones."
RL DNA Res. 4:291-300(1997).
RN (2)
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinzaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB006702; BAB11595.1; -;
DR EMBL: AY062781; AAL32859.1; -;
DR EMBL: AY081645; AAM1207.1; -;
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

Query Match 61.3%; Score 49; DB 10; Length 783;
Best Local Similarity 72.7%; Pred. No. 2.6;
Matches 8; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 3 DODMVVVQVWP 13
DB 760 EDSMVVVQVWP 770

RESULT 3
Q9LOE3 PRELIMINARY; PRT; 570 AA.
ID Q9LOE3;
AC Q9LOE3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE F1504.42.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RA Ecker J.R.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,

```

```

RA Khan S., Kim C., Altafi H., Bel B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howling B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thavert A., Tortum M., Vaysberg M., Yu C., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007887; AAF79360.1; -;
DR InterPro: IPR003340; TF_B3.
DR Pfam: PF02362; B3; 1.
SQ SEQUENCE 570 AA; 65041 MW; E64A987BA713DF7E CRC64;

Query Match 56.2%; Score 45; DB 10; Length 570;
Best Local Similarity 53.8%; Pred. No. 9.9;
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 2 DODMVVVQVWP 14
DB 558 DEDEMLVGDDPWP 570

RESULT 4
Q38832 PRELIMINARY; PRT; 164 AA.
ID Q38832;
AC Q38832;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IAA14 (Fragment).
GN IAA14.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RA MEDLINE=95387393; PubMed=7658471;
RA Abel S., Nguyen M.D., Theologis A.;
RT "The PS-IAA4/5-like family of early auxin-inducible mRNAs in
RT Arabidopsis thaliana."
RL J. Mol. Biol. 251:533-549(1995).
DR EMBL: U18416; AAC49055.1; -;
DR InterPro: IPR003311; AUX_IAA.
DR Pfam: PF02309; AUX_IAA; 1.
DR NON_TER 1
SQ SEQUENCE 164 AA; 18232 MW; D7BB6B71F6DD3574 CRC64;

Query Match 55.0%; Score 44; DB 10; Length 164;
Best Local Similarity 61.5%; Pred. No. 3.9;
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 DODMVVVQVWP 14
DB 120 DGDMLVGVDPWP 132

RESULT 5
Q9C5W8 PRELIMINARY; PRT; 228 AA.
ID Q9C5W8;
AC Q9C5W8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IAA14.
GN D13315C/ATG14550.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.

```


RA	Yamada K, Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
RA	Goldsmith A.D., Lee J.M., Toriumi M., Yu G., Brooks S., Chao Q.,
RA	Chen H., Karlín-Neumann G., Kim C., Lam B., Miranda M., Nguyen M.,
RA	Palm C.J., Shih P., Southwick A., Davis R.W., Ecker J.R.,
RA	Theologis A.;
RT	"Full Length cDNA of gene d13156/AT4G14550 (GI:7268161).";
RL	Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AF34718; AAC50096.1; "
DR	InterPro: IPR003311; AUX_1AA.
DR	Pfam: PF02309; AUX_1AA; 1.
SO	SEQUENCE 228 AA; 25044 MW; A76A9F733538E372 CRC64;
QY	2 DODOMVVQVPM 14
Db	184 DGDMLVGDVPM 196
RESULT 6	
ID	023311
AC	PRELIMINARY; PRT; 234 AA.
OC	023311;
DT	01-JAN-1998 (TREMBlrel. 05, Created)
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT	01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE	IAA7-like protein (IAA7 like protein).
GN	DL315C OR AT4G14550.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC	eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Bevan M., Slekema W., Murphy G., Wambutt R., Pohl T., Terry N.,
RA	Kreis M., Kavanagh T., Entlan K.D., Rieger M., James R.,
RA	Pugdomech P., Hatzopoulos P., Obermayer B., Dusterhoft A.,
RA	Jones G., Palme K., Ansgore W., Delsen M., Bancroft I., Mewes H.W.,
RA	Schneider C., Chalvatzis N.,
RL	Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
RL	[2]
RP	SEQUENCE FROM N.A.
RA	EU Arabidopsis sequencing project.
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL: Z97336; CAB46059.1; "
DR	EMBL: AL161539; CAB78497.1; "
DR	InterPro: IPR003311; AUX_1AA.
DR	Pfam: PF02309; AUX_1AA; 1.
SO	SEQUENCE 234 AA; 25667 MW; 340A94E95B23FB60 CRC64;
Query Match	55.0%; Score 44; DB 10; Length 234;
Best Local Similarity	61.5%; Pred. NO. 5.7;
Matches 8; Conservative	1; Mismatches 4; Indels 0; Gaps 0;
QY	2 DODOMVVQVPM 14
Db	184 DGDMLVGDVPM 196
RESULT 7	
ID	09XIDA
AC	PRELIMINARY; PRT; 619 AA.
OC	09XIDA;
DT	01-NOV-1999 (TREMBlrel. 12, Created)
DT	01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT	01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE	F23M19.4 protein.
GN	F23M19.4.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC	Spermatophyta: Magnoliophyta: eudicotyledons: core eudicots: Rosidae;
OC	eucosids II: Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA:
RA	vyssotskaya V.S., Schwartz J.R., Yu G., Tortumt M., Lenz C., Liu S.,
RA	Lee J., Liu A., Li J., Kremetskaya I., Luros J., Gonzalez A.,
RA	Altai H., Araujo R., Chao Q., Conn L., Conway A.B., Dunn P.,
RA	Hansen N., Hlizer L., Kim C., Palm C., Rowley D., Shinn P., Walker M.,
RA	Davis R.W., Ecker J.R., Federspiel N.A., Theologis A.;
RT	"Arabidopsis thaliana chromosome 1 BAC F2M19 sequence. ";
RL	Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CV. COLUMBIA:
RA	Theologis A.;
RL	Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AC007454; AAD39615.1; -
DR	InterPro: IPR003311; AUX_1AA.
DR	InterPro: IPR003340; TF_B3.
DR	Pfam: PF02309; AUX_1AA; 1.
DR	Pfam: PF02362; B3; 1.
SO	SEQUENCE 619 AA; 69995 MW; 1DA4DD7BC49327E7 CRC64;
Qy	2 DODOMVYQVPP 14
Db	1:1:1:1
Db	592 DEDKMLVGDDPP 604
RESULT 8	
Q9LYN1	PRELIMINARY; PRT; 719 AA.
AC	Q9LYN1;
DT	01-OCT-2000 (TREMblrel. 15, Created)
DT	01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT	01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE	Hypothetical 77.3 kDa protein.
GN	F18021.60.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota: Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC	Spermatophyta; Magnoliophyta; eudicotyledons: core eudicots; Rosidae;
OC	eucosids II: Brassicales; Brassicaceae; Arabidopsids.
OX	NCBI_TaxID=3702;
RN	[1]
RP	SEQUENCE FROM N.A.
RA	Benes V., Wurmbsch E., Drzonek H., Ansoerge W., Mewes H.W., Rudd S.,
RA	Lencke K., Mayer K.F.X., Queller F., Salanoubat M.;
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN	[2]
RP	SEQUENCE FROM N.A.
RA	EU Arabidopsis sequencing project;
RL	Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR	EMBL: AL163763; CAB87409.1; -
DR	InterPro: IPR000719; Euk_kinase.
DR	InterPro: IPR001611; LRR.
DR	InterPro: IPR003592; LRR_out.
DR	Pfam: PF00560; LRR; 7.
DR	Pfam: PF00069; pkinase; 1.
DR	PRINTS: PR00019; LEURICHRP.
DR	ProDom: PD000001; Euk_kinase; 1.
DR	SMART: SM00370; LRR; 6.
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR	ATP-binding; Hypothetical protein; Transferrase.
SO	SEQUENCE 719 AA; 77329 MW; 3B91891A64E0DFD3 CRC64;
Query Match	55.0%; Score 44; DB 10; Length 719;
Best Local Similarity	54.5%; Pred. NO. 19;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;	

QY 3 DQDMVVVQVWP 13
1 : : : 1 : 1 : 1 :
DB 100 QGQVIVITQLPW 110

RESULT 9

Q981L8 PRELIMINARY: PRT: 416 AA.
ID 0981L8
AC 0981L8
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Integrate/recombine.
GN MUR9321.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.,
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP003015; BAB54691.1; -
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR002104; Phage_integrase.
DR Pfam: PF00589; Phage_integrase.1.
DR PROSITE: PS00290; IG_MHC; UNKNOWN.1.
DR Plasmid: Complete proteome.
KW SEQUENCE 416 AA; 46416 MW; 1DAB19A70297A655 CRC64;

Query Match 53.8%; Score 43; DB 16; Length 416;
Best Local Similarity 46.2%; Pred. No. 16;
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 2 DQDMVVVQVWP 14
1 : : : 1 : 1 : 1 :
DB 400 DADRLLAVALPW 412

RESULT 10

Q8UBJ3 PRELIMINARY: PRT: 418 AA.
ID 08UBJ3
AC 08UBJ3
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Intracellular PHB depolymerase.
GN DEPA OR ATU0015 OR AGR_C24.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet M., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmieri A.,
RA Raymond C., Rouse G., Senphimmachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,

RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58.";
RL Science 294:2317-2323(2001).
RN [2]

RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houmel K., Gordon J., Vaudin M., Tatchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Doughty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursan J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL: AE008976; AAL41046.1; -
DR EMBL: AE007944; AAK85840.1; -
KW Complete proteome.
SQ SEQUENCE 418 AA; 46630 MW; 6664887C7A732B34 CRC64;

Query Match 53.8%; Score 43; DB 16; Length 418;
Best Local Similarity 87.5%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 7 VVVOVWP 14
1 : : : 1 : 1 : 1 :
DB 242 VVVOVWP 249

RESULT 11

Q81350 PRELIMINARY: PRT: 423 AA.
ID 081350
AC 081350
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE 5'-adenylisulfate reductase (EC 1.8.99.-).
OS Enteromorpha intestinalis (Hollow green seaweed).
OC Eukaryota; Viridiplantae; Chlorophyta; Ulvophyceae; Ulvales; Ulvaceae;
OC Enteromorpha.
OX NCBI_TaxID=3116;
RN [1]
RP SEQUENCE FROM N.A.
RA Gao Y., Leustek T.;
RT "Cloning of 5'-adenylisulfate (APS) reductase from the marine
RT macroalgae Enteromorpha intestinalis.";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF069951; AAC26855.1; -
DR InterPro: IPR004508; APS_reduc.
DR InterPro: IPR002500; APS_reduc.
DR InterPro: IPR000063; Thiored.
DR Pfam: PF01507; PAPS_reduct.1.
DR Pfam: PF00085; thiored.1.
DR PRINTS: PR00421; THIOREDOXIN.
DR TIGRFAMs: TIGR00424; APS_reduc.1.
KW Oxidoreductase.
SQ SEQUENCE 423 AA; 46808 MW; 808EE4F3BD561F2 CRC64;

Query Match 53.8%; Score 43; DB 10; Length 423;
Best Local Similarity 58.3%; Pred. No. 17;
Matches 7; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 2 DQDMVVVQVWP 13
1 : : : 1 : 1 : 1 :
DB 330 DKDTWVLYLAPW 341

RESULT 12

Q8S3B5 PRELIMINARY: PRT: 462 AA.
ID 08S3B5
AC 08S3B5;

DT 01-JUN-2002 (TREMblrel. 21, Created)
DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Putative glucosyltransferase.
OS Phaseolus lunatus (lima bean) (Phaseolus limensis).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Phaseolus.
OX NCBI_TaxID=3884;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_KINGSTON;
RA Shan X., Martin R.C., Mok D.W.S., Mok M.C.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF489877; AAM09517.1; -
KM Transferase.
RN SEQUENCE 462 AA; 51900 MW; 00C693CDF1E4646E CRC64;
SQ

Query Match 52.5%; Score 42; DB 10; Length 462;
Best Local Similarity 53.8%; Pred. No. 28;
Matches 7; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 2 DODOMVVQVMP 14
DB 10 DKNQVVVLIPEP 22

RESULT 13

ID 039281 PRELIMINARY; PRT; 529 AA.
AC 039281;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Counterpart of HSV-1 gene UL21 and VZV gene 38.
GN 40.
OS Equine herpesvirus 4.
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=10331;
RN [1]
RP SEQUENCE OF 489-529 FROM N.A.
RC STRAIN=NS80567;
RX MEDLINE=90362066; PubMed=2167933;
RA Nicolson L., Cullinane A.A., Onions D.E.;
RT "The nucleotide sequence of an equine herpesvirus 4 gene homologue of the herpes simplex virus 1 glycoprotein H gene.";
RL J. Gen. Virol. 71:1793-1800(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RX MEDLINE=98264497; PubMed=9603335;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-4.";
RL J. Gen. Virol. 79:1197-1203(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RA Telford E.A., Watson M.S., Perry J., Cullinane A.A., Davison A.J.;
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF030027; AAC59558.1; -
DR InterPro: IPR004936; Herpes_UL21.
DR Pfam: PF03252; UL21; 1.
SQ SEQUENCE 529 AA; 58150 MW; E2B737B89D489413 CRC64;

Query Match 52.5%; Score 42; DB 12; Length 529;
Best Local Similarity 45.5%; Pred. No. 32;
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

OY 4 DQMVVQVMP 14
DB 283 DEILVEPMP 293

RESULT 14
ID 09C8N7 PRELIMINARY; PRT; 600 AA.
AC 09C8N7;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Auxin response factor, putative.
GN F7P12.6.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV_COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Etgu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzar L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremetska I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Malt R., Matzall A.,
RA Miltcher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pal G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakono H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uterback T., Van Aken S., Vaysberg M., Vysotska V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis thaliana.";
RL Nature 408:813-820(2000).
DR EMBL: AC023913; AAC51894.1; -
DR InterPro: IPR003311; AUX_1AA.
DR InterPro: IPR003340; TF_B3.
DR Pfam: PF02309; AUX_1AA; 1.
DR Pfam: PF02362; B3; 1.
SQ SEQUENCE 600 AA; 67933 MW; C28A2B4E79073FFF CRC64;

Query Match 52.5%; Score 42; DB 10; Length 600;
Best Local Similarity 53.8%; Pred. No. 37;
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

OY 2 DQDMVVQVMP 14
DB 566 DDDKMLVGDPPMP 578

RESULT 15
ID 09LOE8 PRELIMINARY; PRT; 767 AA.
AC 09LOE8;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE F1504.37.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
[1]
RP SEQUENCE FROM N.A.
RC Ecker J.R.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```
RA Ecker J.R.;
RL Submitted (OCT-1999) to the EMBL/Genbank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Cheuk R., Shin P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
RA Khan S., Kim C., Altari H., Bel B., Chin C., Chlou J., Choi E.,
RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
RA Theologis A., Ecker J.;
RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AC007887; AAF79371.1; -.
DR InterPro; IPR004160; EFTU_Cterm.
DR InterPro; IPR003340; TF_B3.
DR Pfam; PF02362; B3; 1
DR Pfam; PF03143; GTP_EFTU_D3; 1.
SQ SEQUENCE 767 AA; 85951 MW; FE29067F07D2E9F7 CRC64;
```

Query Match 52.5%; Score 42; DB 10; Length 767;
Best Local Similarity 46.2%; Pred. No. 48;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 2 DODQVVVVQVPP 14
:::|:|:| : |||
DB 621 EBDKMLVGEDPMP 633

Search completed: March 27, 2003, 10:06:01
Job time : 9.8536 secs

GenCore version 5.1.4-p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:04:25 ; Search time 3.2439 Seconds
(without alignments)
253.439 Million cell updates/sec

Title: US-09-675-208-3
Perfect score: 80
Sequence: 1 YDQDMVVQVWP 14

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published_Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
- 2: /cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep.*
- 8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep.*
- 10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
- 12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
- 13: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
- 14: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	44	55.0	830	10	US-09-772-134B-14
2	44	55.0	854	9	US-09-754-853A-1103
3	44	55.0	854	9	US-09-754-853A-1105
4	44	55.0	854	9	US-09-754-853A-1111
5	44	55.0	854	9	US-09-754-853A-1113
6	44	55.0	854	9	US-09-754-853A-1115
7	44	55.0	877	9	US-09-754-853A-1104
8	44	55.0	877	9	US-09-754-853A-1110
9	44	55.0	877	9	US-09-754-853A-1112
10	44	55.0	877	9	US-09-754-853A-1114
11	44	55.0	877	9	US-09-754-853A-1118
12	42	52.5	90	10	US-09-864-761-44178
13	40	50.0	46	9	US-09-798-889-50
14	40	50.0	854	9	US-09-754-853A-1098
15	40	50.0	854	9	US-09-754-853A-1107
16	40	50.0	854	9	US-09-754-853A-1109
17	40	50.0	854	9	US-09-754-853A-1109
18	40	50.0	877	9	US-09-754-853A-1097
19	40	50.0	877	9	US-09-754-853A-1100

20	40	50.0	877	9	US-09-754-853A-1106	Sequence 1106, Ap
21	40	50.0	877	9	US-09-754-853A-1108	Sequence 1108, Ap
22	38	47.5	404	10	US-09-908-805B-21	Sequence 21, Appl
23	38	47.5	526	10	US-09-855-154-2	Sequence 2, Appl1
24	38	47.5	574	10	US-10-025-380-1107	Sequence 1107, Ap
25	38	47.5	574	10	US-09-922-217-1107	Sequence 1107, Ap
26	38	47.5	722	9	US-10-029-495-5	Sequence 5, Appl1
27	37	46.2	33	10	US-09-281-717-50	Sequence 50, Appl
28	37	46.2	246	9	US-09-885-827-1	Sequence 1, Appl1
29	37	46.2	294	9	US-09-997-267-4	Sequence 4, Appl1
30	37	46.2	388	9	US-09-997-267-2	Sequence 2, Appl1
31	37	46.2	477	9	US-10-047-412A-10	Sequence 10, Appl
32	37	46.2	477	9	US-10-047-412A-29	Sequence 29, Appl
33	37	46.2	477	10	US-09-923-556-2	Sequence 2, Appl1
34	37	46.2	477	10	US-09-923-556-6	Sequence 6, Appl1
35	37	46.2	477	10	US-09-987-025-2	Sequence 2, Appl1
36	37	46.2	907	9	US-10-008-739A-2	Sequence 2, Appl1
37	37	46.2	978	10	US-09-893-817-4	Sequence 4, Appl1
38	37	46.2	978	10	US-09-893-817-8	Sequence 8, Appl1
39	36	45.0	191	9	US-09-950-933A-66	Sequence 66, Appl
40	36	45.0	229	9	US-09-975-719-265	Sequence 265, App
41	36	45.0	338	10	US-09-815-242-10944	Sequence 10944, A
42	36	45.0	400	9	US-09-943-689A-4	Sequence 4, Appl1
43	36	45.0	402	10	US-09-925-300-1320	Sequence 1320, Ap
44	36	45.0	404	9	US-09-943-689A-2	Sequence 2, Appl1
45	36	45.0	1152	9	US-09-738-626-3614	Sequence 3614, Ap

ALIGNMENTS

RESULT 1
US-09-772-134B-14
Sequence 14, Application US/09772134B
Patent No. US20020144310A1
GENERAL INFORMATION:
APPLICANT: Southern Illinois University
APPLICANT: Lightfoot, David
APPLICANT: Meksem, Khalid
TITLE OF INVENTION: ISOLATED POLYNUCLEOTIDES AND POLYPEPTIDES RELATING TO LOCI UND
TITLE OF INVENTION: RESISTANCE TO SOYBEAN CYST NEMATODE AND SOYBEAN SUDDEN DEATH
FILE REFERENCE: 1268/4/2
CURRENT APPLICATION NUMBER: US/09/772,134B
CURRENT FILING DATE: 2001-01-29
PRIOR APPLICATION NUMBER: 60/178,811
PRIOR FILING DATE: 2000-01-28
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14
LENGTH: 830
TYPE: PRT
ORGANISM: soybean
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(830)
OTHER INFORMATION: xaa is any amino acid
US-09-772-134B-14
Query Match 55.0%; Score 44; DB 10; Length 830;
Best Local Similarity 54.5%; Pred. No. 10;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 3 QDQMVVQVWP 13
DB 24 QGVIVQLPW 34
RESULT 2
US-09-754-853A-1103
Sequence 1103, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:

```

1  APPLICANT: Hauge, Brian M.
2  APPLICANT: Parnell, Laurence D.
3  APPLICANT: Parsons, Jeremy D.
4  APPLICANT: Wang, Ming Li
5  TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
6  TITLE OF INVENTION: Soybean Cyst Nematode Resistance
7  FILE REFERENCE: 38-10(15810)B
8  CURRENT APPLICATION NUMBER: US/09/754, 853A
9  CURRENT FILING DATE: 2001-01-05
10 PRIOR APPLICATION NUMBER: US 60/174, 880
11 PRIOR FILING DATE: 2000-01-07
12 NUMBER OF SEQ ID NOS: 1119
13 SEQ ID NO 1103
14 LENGTH: 854
15 TYPE: PRF
16 ORGANISM: Glycine max
17 FEATURE:
18 OTHER INFORMATION: Clone ID: rhy1_peeking_amp11con
19 US-09-754-853A-1103

```

Query Match	55.08;	Score 44;	DB 9;	Length 854;
Best Local Similarity	54.58;	Pred. No. 10;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;

```
QY      3 QDQMNVNVQVPW 13
          | | : | : | |
Db     115 QGQVIVIQLPW 125
```

RESULT 3
US-09-754-853A-1105
; Sequence 1105, Application US/09754853A
; Publication No. US20030005491A1

GENERAL INFORMATION:
 APPLICANT: Hauge, Brian M.
 APPLICANT: Parnell, Laurence D.
 APPLICANT: Parsons, Jeremy D.
 APPLICANT: Wang, Ming Li
 TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
 TITLE OF INVENTION: Soybean Cyst Nematode Resistance
 FILE REFERENCE: 38-10(115810)B
 CURRENT APPLICATION NUMBER: US/09/754,853A
 CURRENT FILING DATE: 2001-01-05
 PRIOR APPLICATION NUMBER: US 60/174,880
 PRIOR FILING DATE: 2000-01-07
 NUMBER OF SEQ ID NOS: 1119
 SEQ ID NO 1105
 LENGTH: 854
 TYPE: PRT
 ORGANISM: Glycine max
 FEATURE:
 OTHER INFORMATION: Clone ID: rhgl_1toyoosuzu_amlp1con
 US-09-754-853A-1105

Query Match	55.0%;	Score 44;	DB 9;	Length 854;
Best Local Similarity	54.5%;	Pred. No. 10;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;

```
QY      3 QDQMNVNVQVPM 13
          | | : | : | | |
Db      115 QGQVIVIQLPW 125
```

```

RESULT 4
US-09-754-853A-1111
; Sequence 1111, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With

```

```

: TITLE OF INVENTION: Soybean Cyst Nematode Resistance
: FILE REFERENCE: 38-10(13810)B
: CURRENT APPLICATION NUMBER: US/09/754,853A
: CURRENT FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 60/174,880
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 1119
: SEQ ID NO 1111
: LENGTH: 854
: TYPE: PRT
: ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Clone ID: rhg1_noir_amplicon
US-09-754-853A-1111

```

Query Match	55.0%;	Score 44;	DB 9;	Length 854;
Best Local Similarity	54.5%;	Pred. No. 10;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;

QY 3 QDQMNVVQVPW 13
 | : : : : :
Db 115 QGQVIVIQLPW 125

```

RESULT 5
US-09-754-853A-1113
: Sequence 1113, Application US/09754853A
: Publication No. US20030005491A1
: GENERAL INFORMATION:
: APPLICANT: Hauge, Brian M.
: APPLICANT: Parnell, Laurence D.
: APPLICANT: Parsons, Jeremy D.
: APPLICANT: Wang, Ming Li
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Soybean Cyst Nematode Resistance
: FILE REFERENCE: 38-10(15810)B
: CURRENT APPLICATION NUMBER: US/09/754, 853A
: CURRENT FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 60/174, 880
: PRIOR FILING DATE: 2000-01-07
: NUMBER OF SEQ ID NOS: 1119
: SEQ ID NO 1113
: LENGTH: 854
: TYPE: PRT
: ORGANISM: Glycine max
: FEATURE:
: OTHER INFORMATION: Clone ID: rhg1_lee_amplicon
US-09-754-853A-1113

```

Query Match	55.0%;	Score 44;	DB 9;	Length 854;
Best Local Similarity	54.5%;	Pred. No. 10;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;

```
QY      3 QDQMNVVQVPW 13
          | |::|:|:| |
Db      115 QGQVIVIQLPW 125
```

```

RESULT 6
US-09-754-853A-1115
: Sequence 1115, Application US/09754853A
: Publication No. US20030005491A1
: GENERAL INFORMATION:
: APPLICANT: Hauge, Brian M.
: APPLICANT: Parnell, Laurence D.
: APPLICANT: Parsons, Jeremy D.
: APPLICANT: Wang, Ming Li
: TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
: TITLE OF INVENTION: Soybean Cyst Nematode Resistance
: FILE REFERENCE: 38-10(15810)B
: CURRENT APPLICATION NUMBER: US/09/754.853A
: CURRENT FILING DATE: 2001-01-05
: PRIOR APPLICATION NUMBER: US 60/174,880

```

```
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1115
; LENGTH: 854
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_p1200499_amplicon
US-09-754-853A-1115

Query Match
Best Local Similarity 54.5%; Score 44; DB 9; Length 854;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVQVPM 13
|:::|:|:|
Db 138 OGVIVIQLPW 148

RESULT 7
US-09-754-853A-1102
; Sequence 1102, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1102
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_pekling_amplicon
US-09-754-853A-1102

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 877;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVQVPM 13
|:::|:|:|
Db 138 OGVIVIQLPW 148

RESULT 8
US-09-754-853A-1104
; Sequence 1104, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1104
; LENGTH: 877
; TYPE: PRT
```

```
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_toyosuzu_amplicon
US-09-754-853A-1104

Query Match
Best Local Similarity 54.5%; Score 44; DB 9; Length 877;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVQVPM 13
|:::~|:|:|
Db 138 OGVIVIQLPW 148

RESULT 9
US-09-754-853A-1110
; Sequence 1110, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1110
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_nolr_amplicon
US-09-754-853A-1110

Query Match
Best Local Similarity 55.0%; Score 44; DB 9; Length 877;
Matches 6; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 3 ODOMVVVQVPM 13
|:::~|:|:|
Db 138 OGVIVIQLPW 148

RESULT 10
US-09-754-853A-1112
; Sequence 1112, Application US/09754853A
; Publication No. US20030005491A1
; GENERAL INFORMATION:
; APPLICANT: Hauge, Brian M.
; APPLICANT: Parnell, Laurence D.
; APPLICANT: Parsons, Jeremy D.
; APPLICANT: Wang, Ming Li
; TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
; FILE REFERENCE: 38-10(15810)B
; CURRENT APPLICATION NUMBER: US/09/754,853A
; PRIOR FILING DATE: 2001-01-05
; PRIOR APPLICATION NUMBER: US 60/174,880
; PRIOR FILING DATE: 2000-01-07
; NUMBER OF SEQ ID NOS: 1119
; SEQ ID NO 1112
; LENGTH: 877
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: rhg1_lee_amplicon
US-09-754-853A-1112
```

Query Match	55.08;	Score 44;	DB 9;	Length 877;
Best Local Similarity	54.58;	Pred. No. 11;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0

```
QY      3 QDQMNVNQPW 13
          | |:|:|:|
Db     138 QGQVIVIQLPW 148
```

RESULT 11
US-09-754

Sequence 1114, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parnell, Laurence D.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
CURRENT FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 60/174,880
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 1114
LENGTH: 877
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: rhtg1_pi200499_amlp1con
US-09-754-853A-1114

Query Match	55.0%;	Score 44;	DB 9;	Length 877;
Best Local Similarity	54.5%;	Pred. No. 11;		
Matches	6;	Conservative	4;	Mismatches 1;
				Indels 0;
				Gaps 0;

QY 3 QDQMNVVQVPM 13
1 1::1:1:1
Db 138 QGQVIVIQLPW 148

RESULT 12
US-09-864

US-09-864-761-44178
Sequence 44178, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeonica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006653
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/006655
 PRIOR FILING DATE: 2001-01-30
 PRIOR APPLICATION NUMBER: PCT/US01/006668
 PRIOR FILING DATE: 2001-01-30

```

; OTHER INFORMATION: MAP TO AC005866.3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.7
; OTHER INFORMATION: EST_HUMAN HIT: AW962016.1, EVALUE 2.00e-23
; OS-09-864 -761 -44178
```

Query Match	52.58;	Score 42;	DB 10;	Length 90;
Best Local Similarity	60.08;	Pred. No. 2;		
Matches	6;	Conservative	2;	Mismatches 2;
			Indels	0;
			Gaps	0;

QY	5	QMVVVQVPWP	14
		: :	
Db	47	QLVIVNFPWP	56

RESULT 13
US-09-798

```

Sequence 0: Application US/09798889
Publication NO. US20030004324A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: 31 Human secreted proteins
FILE REFERENCE: P2026p1
CURRENT APPLICATION NUMBER: US/09/798, 889
CURRENT FILING DATE: 2001-03-06
PRIORITY APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/393,022
PRIORITY FILING DATE: EARLIER FILING DATE: 1999-09-09
PRIORITY APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,714
PRIORITY FILING DATE: EARLIER FILING DATE: 1998-03-12
PRIORITY APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686
PRIORITY FILING DATE: EARLIER FILING DATE: 1998-03-12
PRIORITY APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,686
PRIORITY FILING DATE: EARLIER FILING DATE: 1998-03-12
PRIORITY APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/077,696
PRIORITY FILING DATE: EARLIER FILING DATE: 1998-03-12
NUMBER OF SEQ ID NOS: 185
SOFTWARE: patentIn Ver. 2.0
SEQ ID NO 50
LENGTH: 46
TYPE: PRT
ORGANISM: Homo sapiens

```


FEATURE:
NAME/KEY: SITE
LOCATION: (46)
OTHER INFORMATION: Xaa equals stop translation
US-09-798-889-50

Query Match 50.0%; Score 40; DB 9; Length 46;
Best Local Similarity 66.7%; Pred. No. 2.1;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 6 MYVVOVPM 14
: : : : :
: : : : :
DB 18 LVVVSLLPM 26

RESULT 14
US-09-754-853A-1098
Sequence 1098, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Parnell, Laurence D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
PRIOR APPLICATION NUMBER: 2001-01-05
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 1098
LENGTH: 854
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: 240017_region_G3
US-09-754-853A-1098

Query Match 50.0%; Score 40; DB 9; Length 854;
Best Local Similarity 55.6%; Pred. No. 52;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 5 QMYYVOVPM 13
: : : : :
: : : : :
DB 117 QVIVIQLPW 125

RESULT 15
US-09-754-853A-1101
Sequence 1101, Application US/09754853A
Publication No. US20030005491A1
GENERAL INFORMATION:
APPLICANT: Hauge, Brian M.
APPLICANT: Parnell, Laurence D.
APPLICANT: Parsons, Jeremy D.
APPLICANT: Wang, Ming Li
TITLE OF INVENTION: Nucleic Acid Molecules And Other Molecules Associated With
FILE REFERENCE: 38-10(15810)B
CURRENT APPLICATION NUMBER: US/09/754,853A
CURRENT FILING DATE: 2001-01-05
PRIOR APPLICATION NUMBER: US 60/174,880
PRIOR FILING DATE: 2000-01-07
NUMBER OF SEQ ID NOS: 1119
SEQ ID NO 1101
LENGTH: 854
TYPE: PRT
ORGANISM: Glycine max
FEATURE:
OTHER INFORMATION: Clone ID: rhg1_A3244_amplicon
US-09-754-853A-1101

Query Match 50.0%; Score 40; DB 9; Length 854;
Best Local Similarity 55.6%; Pred. No. 52;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 5 QMYYVOVPM 13
: : : : :
: : : : :
DB 117 QVIVIQLPW 125

Search completed: March 27, 2003, 10:20:37
Job time : 4.2439 secs

GenCore version 5.1.4_p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 4.63415 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-2
105
Sequence: 1 PVSVCFFVGFDAEPDSRH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	19	4	US-08-846-234-2
2	105	100.0	784	4	US-08-846-234-5
3	42	40.0	22	4	US-08-905-223-460
4	41	39.0	212	4	US-08-861-774E-22
5	41	39.0	212	4	US-08-861-774E-34
6	40	38.1	291	4	US-09-105-697-4
7	40	38.1	291	4	US-09-105-697-5
8	40	38.1	291	4	US-09-105-697-6
9	40	38.1	834	1	US-07-977-434-8
10	40	38.1	834	1	US-07-977-434-10
11	40	38.1	834	1	US-08-073-384C-6
12	40	38.1	834	1	US-08-254-359A-6
13	40	38.1	834	1	US-08-384-490-31
14	40	38.1	834	1	US-08-483-043-6
15	40	38.1	834	1	US-08-459-363-31
16	40	38.1	834	1	US-08-458-819-8
17	40	38.1	834	1	US-08-458-819-10
18	40	38.1	834	1	US-08-481-238-6
19	40	38.1	834	2	US-08-471-066B-6
20	40	38.1	834	2	US-08-484-956-6
21	40	38.1	834	2	US-08-757-653-6
22	40	38.1	834	2	US-08-599-491-6
23	40	38.1	834	2	US-08-756-386-6
24	40	38.1	834	2	US-08-823-516-6
25	40	38.1	834	3	US-08-682-853A-6
26	40	38.1	834	3	US-08-759-038-6
27	40	38.1	834	3	US-08-758-314-6

28	40	38.1	834	4	US-09-350-309-6	Sequence 6, Appl1
29	40	38.1	834	4	US-08-520-946-6	Sequence 6, Appl1
30	40	38.1	834	5	PCT-US91-07035-8	Sequence 8, Appl1
31	40	38.1	834	5	PCT-US91-07035-10	Sequence 10, Appl1
32	40	38.1	834	5	PCT-US95-14418-2	Sequence 2, Appl1
33	40	38.1	834	5	PCT-US95-15327-2	Sequence 2, Appl1
34	39.5	37.6	435	2	US-08-338-530A-2	Sequence 2, Appl1
35	39.5	37.6	435	2	US-09-267-384-2	Sequence 2, Appl1
36	39	37.1	266	2	US-07-857-224B-39	Sequence 39, Appl1
37	39	37.1	481	4	US-08-843-659-4	Sequence 4, Appl1
38	39	37.1	560	5	PCT-US95-14418-5	Sequence 5, Appl1
39	39	37.1	560	5	PCT-US95-15327-5	Sequence 5, Appl1
40	39	37.1	597	5	PCT-US95-14418-4	Sequence 4, Appl1
41	39	37.1	597	5	PCT-US95-15327-4	Sequence 4, Appl1
42	39	37.1	833	1	US-08-073-384C-8	Sequence 8, Appl1
43	39	37.1	833	1	US-08-254-359A-8	Sequence 8, Appl1
44	39	37.1	833	1	US-08-483-043-8	Sequence 8, Appl1
45	39	37.1	833	1	US-08-481-238-8	Sequence 8, Appl1

ALIGNMENTS

RESULT 1
US-08-846-234-2
; Sequence 2, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chleko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESS: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846, 234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 19 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FRAGMENT TYPE: Internal
; US-08-846-234-2

Query Match 100.0%; Score 105; DB 4; Length 19;
Best Local Similarity 100.0%; Pred. No. 2e-10;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVSVCFFVGFDAEPDSRH 19
DB 1 PVSVCFFVGFDAEPDSRH 19

```

RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P. C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-234-5

Query Match          100.0%; Score 105; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 1,2e-08;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 PVSVCFCVGFDSSEPSRH 19
Db 61 PVSVCFCVGFDSSEPSRH 79

RESULT 3
US-08-905-223-460
; Sequence 460, Application US/08905223
; Patent No. 6222029
; GENERAL INFORMATION:
; APPLICANT: Edwards, Jean-Baptiste D.
; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: 5' ESTS FOR SECRETED PROTEINS
; NUMBER OF SEQUENCES: 503
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 501 West Broadway
; CITY: San Diego
; STATE: California
; COUNTRY: USA
; ZIP: 92101-3505
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: Win95
; SOFTWARE: Word
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/905,223

```

```

; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 460:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 22 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: LINEAR
; MOLECULE TYPE: PROTEIN
; ORIGINAL SOURCE:
; ORGANISM: Homo Sapiens
; TISSUE TYPE: Brain
; FEATURE:
; NAME/KEY: sig-peptide
; LOCATION: -17..-1
; IDENTIFICATION METHOD: Von Heijne matrix
; OTHER INFORMATION: score 5.2
; OTHER INFORMATION: seq IMCLIGLKANASS/ET
US-08-905-223-460

Query Match          40.0%; Score 42; DB 4; Length 22;
Best Local Similarity 36.8%; Pred. No. 2.1;
Matches 7; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

QY 1 PVSVCFCVGFDSSEPSRH 19
Db 2 PVSIMCLIGLKANASSETH 20

RESULT 4
US-08-861-774E-22
; Sequence 22, Application US/08861774E
; Patent No. 6297007
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/08/861,774E
; CURRENT FILING DATE: 1997-05-22
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 22
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Xanthoparmelia cumberlandia
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa-unknown amino acid
US-08-861-774E-22

Query Match          39.0%; Score 41; DB 4; Length 212;
Best Local Similarity 43.8%; Pred. No. 37;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCFCVGFDSSEPD 16
Db 22 PEDVGCYIGACATDYD 37

RESULT 5
US-08-861-774E-34
; Sequence 34, Application US/08861774E

```

```

: Patent No. 6297007
: GENERAL INFORMATION:
: APPLICANT: Waters, Barbara
: APPLICANT: Miao, Vivian
: APPLICANT: Ho, Yap
: APPLICANT: Tong, Seow
: TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
: TITLE OF INVENTION: BIOACTIVE MOLECULES
: FILE REFERENCE: 9993-006
: CURRENT APPLICATION NUMBER: US/08/861,774E
: CURRENT FILING DATE: 1997-05-22
: NUMBER OF SEQ ID NOS: 94
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 34
: LENGTH: 212
: TYPE: PRF
: ORGANISM: Leptogium corniculatum
: US-08-861-774E-34

```

Query Match	39.08;	Score 41;	DB 4;	Length 212;
Best Local Similarity	43.80;	Pred. NO. 37;		
Matches	7;	Conservative	4;	Mismatches 5;
				Indels 0;
				Gaps 0;

```

Oy      1 PVSVCFFVGFDASEPD 16
          | |||::| |::|
Db      22 PEDVGCYTGACATDYD 37

```

RESULT 6
US-09-105-697-4

```

1  GENERAL INFORMATION:
2  APPLICANT: Gelfand Ph.D., David H.
3  APPLICANT: Reichert, Fred L.
4  TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASE
5  NUMBER OF SEQUENCES: 11
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: Roche Molecular Systems
8  STREET: 1080 U.S. Highway 202
9  CITY: Branchburg
10 STATE: New Jersey
11 COUNTRY: United States
12 ZIP: 08876
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: PatentIn Release #1.0, Version #1.25
19 CURRENT APPLICATION DATA:
20 APPLICATION NUMBER: US/09/105,697

```

```
QY      7 EVGFDASEPPDSRH 19
        |||||
Db      64 EVVEDAKAPSERH 76
```

RESULT 7
 US-09-105-697-5
 : Sequence 5, Application US/09105697
 : Patent No. 6228628
 : GENERAL INFORMATION:
 : APPLICANT: Gelland Ph.D., David H.
 : APPLICANT: Reichert, Fred L.
 : TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASED
 : NUMBER OF SEQUENCES: 11
 : CORRESPONDENCE ADDRESSES:
 : ADDRESSSEE: Roche Molecular Systems
 : STREET: 1080 U.S. Highway 202
 : CITY: Branchburg
 : STATE: New Jersey
 : COUNTRY: United States
 : ZIP: 08876
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: PatentIn Release #1.0, Version #1.25
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/09/105,697
 : FILING DATE:
 :

```
QY 7 FVGFDASEPDSRH 19
    |||||
Db 64 FVFEDAKAPSFRR 76
```

```

1  RESULT 8
2  US-09-105-697-6
3  : Sequence 6, Application US/09105697
4  : Patent No. 6228628
5  :
6  : GENERAL INFORMATION:
7  :
8  : APPLICANT: Gelland Ph.D., David H.
9  : APPLICANT: Relchert, Fred L.
10 :
11 : TITLE OF INVENTION: MUTANT CHIMERIC DNA POLYMERASE
12 :
13 : NUMBER OF SEQUENCES: 11
14 :
15 : CORRESPONDENCE ADDRESS:
16 : ADDRESS: Roche Molecular Systems
17 : STREET: 1080 U.S. Highway 202
18 :
19 : CITY: Branchburg
20 :
21 : STATE: New Jersey
22 :
23 : COUNTRY: United States
24 :
25 : ZIP: 08876
26 :
27 : COMPUTER READABLE FORM:
28 :
29 : MEDIUM TYPE: Floppy disk
30 :

```

Query Match	38.1%	Score 40:	DB 4:	Length 291:
Best Local Similarity	61.5%	Pred. No. 75:		
Matches	8:	Conservative	5:	Indels 0:
		Mismatches	5:	Gaps 0:

```

;
;      country:  United States
;      zip:      08876
;
;      computer readable form:
;      medium type:  floppy disk
;

```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/105,697
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pcty Ph.D., Douglas A.
REGISTRATION NUMBER: 35321
REFERENCE/DOCKET NUMBER: 1043
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510)814-2974
TELEFAX: (510)814-2977
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 291 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-105-697-6

Query Match 38.1%; Score 40; DB 4; Length 291;
Best local Similarity 61.5%; Pred. No. 75;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 7 FVGFDAEPDSRH 19
DB 64 FVFEAKAPSRH 76

RESULT 9
US-07-977-434-8
Sequence 8, Application US/07977434
Patent No. 5466591
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Adrason, Richard D.
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 7
SOFTWARE: Wordperfect 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/977,434
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,490
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,466
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,213
FILING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,394
FILING DATE: 15-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 143,441
FILING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 063,509
FILING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FILING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO/PCT/US90/07641
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FILING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FILING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FILING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FILING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Liann Casert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-977-434-8

Query Match 38.1%; Score 40; DB 1; Length 834;
Best local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
DB 64 FVFEAKAPSRH 76

RESULT 10
US-07-977-434-10
Sequence 10, Application US/07977434
Patent No. 5466591
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Adrason, Richard D.
TITLE OF INVENTION: 5' TO 3' EXONUCLEASE MUTATIONS OF
TITLE OF INVENTION: THERMOSTABLE DNA POLYMERASES
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
ZIP: 07110-1199
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: 7
SOFTWARE: Wordperfect 2.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/977,434
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,490

FLING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,466
FLING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 590,213
FLING DATE: 28-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 523,394
FLING DATE: 15-MAY-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 143,441
FLING DATE: 12-JAN-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 063,509
FLING DATE: 17-JUN-1987
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 899,241
FLING DATE: 22-AUG-1986
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 746,121
FLING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US90/07641
FLING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 585,471
FLING DATE: 20-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 455,611
FLING DATE: 22-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 609,157
FLING DATE: 02-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 557,517
FLING DATE: 24-JUL-1990
ATTORNEY/AGENT INFORMATION:
NAME: Luann Caert
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: Case No. 5466591 8753
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2972
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-977-434-10

Query Match 38.1% Score 40: DB 1: Length 834:
Best Local Similarity 61.5% Pred. No. 2.4e+02;
Matches 8: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDASEPDSRH 19
DB 64 FVFDKAKPSFRH 76

RESULT 11
US-08-073-384C-6
Sequence 6, Application US/08073384C
Patent No. 5614311
GENERAL INFORMATION:
APPLICANT: Dahlberg, James E.
APPLICANT: Lyamichay, Victor I.
APPLICANT: Brow, Mary Ann D.
TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/073,384C
FLING DATE: 04-JUN-1993
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/986,330
FLING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-00613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-073-384C-6

Query Match 38.1% Score 40: DB 1: Length 834:
Best Local Similarity 61.5% Pred. No. 2.4e+02;
Matches 8: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDASEPDSRH 19
DB 64 FVFDKAKPSFRH 76

RESULT 12
US-08-254-359A-6
Sequence 6, Application US/08254359A
Patent No. 5614402
GENERAL INFORMATION:
APPLICANT: DAHLBERG, JAMES E.
APPLICANT: LYAMICHEV, VICTOR I.
APPLICANT: BROW, MARY ANN D.
TITLE OF INVENTION: 5' NUCLEASES DERIVED FROM THERMOSTABLE
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERSTOCK, MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/254,359A
FLING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FLING DATE: 06-JUN-1993
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-234-359A-6

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAPSRH 76

RESULT 13
US-08-384-490-31
Sequence 31, Application US/08384490
Patent No. 5618711
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.
APPLICANT: Scofield, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermophilus DNA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/384,490
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/148,133
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Stas, Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: 8887
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-384-490-31

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAPSRH 76

RESULT 14
US-08-483-043-6
Sequence 6, Application US/08483043
Patent No. 5691142
GENERAL INFORMATION:
APPLICANT: Dahlberg, James E.
APPLICANT: Lyamichev, Victor I.
APPLICANT: Brow, Mary Ann D.
TITLE OF INVENTION: SYNTHESIS-DEFICIENT THERMOSTABLE DNA
TITLE OF INVENTION: POLYMERASE
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: HAVERTOCK, MEDLEN & CARROLL
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/483,043
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/073,384
FILING DATE: 04-JUN-1993
APPLICATION NUMBER: US 07/986,330
FILING DATE: 07-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Carroll, Peter G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-00613
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/705-8410
TELEFAX: 415/397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-483-043-6

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
||| |
Db 64 FVFDKAPSRH 76

RESULT 15
US-08-459-383-31
Sequence 31, Application US/08459383
Patent No. 5741690
GENERAL INFORMATION:
APPLICANT: Gelfand, David H.

APPLICANT: Lawyer, Frances C.
APPLICANT: Scofield, Susanne
TITLE OF INVENTION: Recombinant Expression Vectors and
TITLE OF INVENTION: Purification Methods for Thermus Thermophilus DNA
TITLE OF INVENTION: Polymerase
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hoffmann-La Roche Inc.
STREET: 340 Kingsland Street
CITY: Nutley
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07110
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,383
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/384,490
FILING DATE:
APPLICATION NUMBER: US/08/148,133
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Sias, Stacey R.
REGISTRATION NUMBER: 32,630
REFERENCE/DOCKET NUMBER: 8867
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 814-2863
TELEFAX: (510) 814-2977
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
US-08-459-383-31

Query Match 38.1%; Score 40; DB 1; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.4e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
DB 64 FVVFDAKAPSPRH 76

Search completed: March 27, 2003, 10:07:34
Job time : 5.63415 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 13.439 Seconds
(without alignments)
188.389 Million cell updates/sec

Title: US-09-675-208-2

Perfect score: 105

Sequence: 1 PVSVCFFVGFDPASEPDSRH 19

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_101002:*

- 1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
- 2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
- 3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
- 4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
- 5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
- 6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
- 7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
- 8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
- 9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
- 10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
- 11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
- 12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
- 13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
- 14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
- 15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
- 16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
- 17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
- 18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
- 19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
- 20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
- 21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
- 22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
- 23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	105	100.0	19	AAW53568	Cucumber raffinose
2	105	100.0	19	AAV17423	Cucumber raffinose
3	105	100.0	784	AAW53570	Cucumber raffinose
4	105	100.0	784	AAV17417	Cucumber raffinose
5	71	67.6	780	AAW53569	Soybean protein: S
6	71	67.6	781	AAW57887	Soybean raffinose
7	71	67.6	781	AAW57887	Amino acid sequenc
8	71	67.6	781	AAW57887	Soybean raffinose
9	63	60.0	799	AAW57886	Broad bean raffin
10	63	60.0	799	AAV30142	Amino acid sequenc

11	62	59.0	758	21	AAV70978
12	61	58.1	783	20	AAV32073
13	47.5	45.2	770	21	AAV70976
14	47	44.8	775	21	AAW15865
15	47	44.8	253	21	AAW15864
16	47	44.8	295	21	AAW15863
17	46.5	44.3	777	20	AAV32074
18	46.5	44.3	783	23	ABW93664
19	45	42.9	763	21	AAV70977
20	44	41.9	112	18	AAW08433
21	44	41.9	122	22	ABW63148
22	44	41.9	428	22	ABW60476
23	43	41.0	86	23	ABP31035
24	43	41.0	313	22	AAW50675
25	43	41.0	441	20	AAV40501
26	43	41.0	447	21	AAW51496
27	42.5	40.5	841	21	AAV70981
28	42	40.0	22	20	AAV13176
29	42	40.0	257	17	AAW94460
30	42	40.0	382	22	AAW93040
31	42	40.0	382	22	AAW76725
32	41	39.0	212	20	AAW82678
33	41	39.0	212	20	AAW82678
34	41	39.0	212	20	AAW82685
35	41	39.0	823	22	ABW19520
36	40	38.1	48	22	AAO10988
37	40	38.1	52	22	ABG01020
38	40	38.1	52	22	ABG01021
39	40	38.1	78	22	AAU47171
40	40	38.1	129	22	AAW00853
41	40	38.1	144	20	AAV59807
42	40	38.1	145	20	AAV48481
43	40	38.1	146	22	ABG20734
44	40	38.1	232	22	AAW82256
45	40	38.1	283	23	ABW57205

ALIGNMENTS

RESULT 1
ID AAW53568 standard; peptide; 19 AA.
XX
AC AAW53568;
XX
DT 06-JUL-1998 (first entry)
XX
DE Cucumber raffinose synthase residues 61 to 79.
XX
XX Cucumber; raffinose synthase; sucrose; galactinol.
XX
OS Cucumis sativus.
XX
PN JPI0084973-A.
XX
PD 07-APR-1998.
XX
PE 28-APR-1997; 97JP-011124.
XX
PR 26-JUL-1996; 96JP-0198079.
XX
PR 26-APR-1996; 96JP-0107682.
XX
PA (AJIN) AJINOMOTO KK.
XX
DR WPI; 1998-264858/24.
XX
PT Raffinose synthase gene - useful for preparation of raffinose in
XX transformed plant
PS Example 2; Page 17; 26pp: Japanese.
XX
XX The present sequence is a cucumber raffinose synthase fragment.

Soybean raffinose
Sugarcane raffinose
Rice raffinose syn
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Mustard raffinose
Herbicide raffinose
Rice raffinose syn
Agaricus bisporus
Drosophila melanog
Drosophila melanog
Human hydrolyase-11
C. elegans I-beta
A. thaliana yael p
Arabidopsis thalia
Wheat raffinose sy
Human secreted pro
Uroporphyrinogen I
C glutamicum prote
Corynebacterium gl
X. cumberlania ty
L. corniculatum ty
Novel human diago
Human polypeptide
Novel human diago
Novel human diago
Propionibacterium
Human bone marrow
Human normal ovar
Human breast tumo
S. epidermidis ope
Mouse ischaemic co

CC Raffinose synthase forms raffinose from sucrose and galactinol, has
 CC an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees
 CC C, has a molecular weight of 75 to 95 kDa by gel filtration or 90
 CC to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is
 CC inhibited by Iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 19 AA;

Query Match 100.0%; Score 105; DB 19; Length 19;
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDASEPDSSRH 19
 |||||
 DB 1 PVSVCFCVGFDASEPDSSRH 19

RESULT 2

AAV17423
 ID AAV17423 standard; peptide; 19 AA.

XX AAV17423;

XX 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase peptide SEQ ID NO:2.

XX Raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

XX JP11123080-A.

XX 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

DR WPI; 1999-340516/29.

PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

PS Example 2; Page 21; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents a raffinose
 CC synthase peptide from cucumber.

XX Sequence 19 AA;

Query Match 100.0%; Score 105; DB 20; Length 19;
 Best Local Similarity 100.0%; Pred. No. 2.6e-10;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDASEPDSSRH 19
 |||||
 DB 1 PVSVCFCVGFDASEPDSSRH 19

RESULT 3

AAW53570
 ID AAW53570 standard; Protein; 784 AA.

XX AAW53570;

XX 06-JUL-1998 (first entry)

XX Cucumber raffinose synthase.

XX Cucumber: raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

XX JP10084973-A.

XX 07-APR-1998.

PF 28-APR-1997; 97JP-0111124.

PR 26-JUL-1996; 96JP-0198079.

PR 26-APR-1996; 96JP-0107682.

XX (AJIN) AJINOMOTO KK.

DR WPI; 1998-264858/24.

DR N-PSDB; AAV22250.

PT Raffinose synthase gene - useful for preparation of raffinose in
 PT transformed plant

PS Claim 3; Pages 17-20; 26pp; Japanese.

CC The present sequence is cucumber raffinose synthase, which
 CC forms raffinose from sucrose and galactinol, has an optimum pH of
 CC 6 to 8 and working temperature of 35 to 40 degrees C, has a
 CC molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa
 CC by PAGE and SDS-PAGE under reductive conditions and is inhibited by
 CC Iodoacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA;

Query Match 100.0%; Score 105; DB 19; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFCVGFDASEPDSSRH 19
 |||||
 DB 61 PVSVCFCVGFDASEPDSSRH 79

RESULT 4

AAV17417
 ID AAV17417 standard; Protein; 784 AA.

XX AAV17417;

XX 29-JUL-1999 (first entry)

DE Cucumber raffinose synthase.

XX Raffinose synthase; sucrose; galactinol.

XX Cucumis sativus.

XX JP11123080-A.

XX 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

XX (AJIN) AJINOMOTO KK.

DR WPI; 1999-340516/29.

DR N-PSDB; AAX61238.

PT New raffinose synthase gene - for production of raffinose from
 PT sucrose and galactinol

XX Claim 2; Page 25-27; 37pp; Japanese.
 PS
 CC The present invention describes a raffinose synthase, having an activity
 CC of forming raffinose from sucrose and galactinol. The raffinose synthase
 CC gene can be used for expression in a plant for the production of
 CC raffinose. The raffinose synthase can give raffinose from sucrose
 CC and galactinol efficiently. The present sequence represents raffinose
 CC synthase from cucumber.
 XX
 SQ Sequence 784 AA;

Query Match 100.0%; Score 105; DB 20; Length 784;
 Best Local Similarity 100.0%; Pred. No. 1.2e-08;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 PVSVCFGFSDASEPDSRH 19
 |||||
 DB 61 PVSVCFGFSDASEPDSRH 79

RESULT 5

AAB98659
 ID AAB98659 standard; protein; 780 AA.

XX
 AC AAB98659;

DT 17-AUG-2001 (first entry)

DE Soybean protein; SEQ ID 1.

KW Mutant; mutein; raffinose synthase; raffinose oligosaccharide reduction;
 plant; soybean.

XX
 OS Glycine max.

PN JP2001078783-A.

PD 27-MAR-2001.

PF 03-JUL-2000; 2000JP-0200571.

PR 09-JUL-1999; 99JP-0196036.

PA (SUMO) SUMITOMO CHEM CO LTD.

DR WPI: 2001-313373/33.

DR N-PSDB: AAH27438.

PT Novel mutant protein of raffinose synthase is useful for reducing the
 raffinose oligosaccharide content in a plant body -

PS Disclosure: Page 18-20; 30pp; Japanese.

XX
 CC The present invention relates to a mutant protein of raffinose synthase
 CC in which at least one aromatic amino acid present at the position of
 CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
 CC mutant protein can be used for reducing the raffinose oligosaccharide
 CC content in a plant body. The present protein from soybean, was used in
 CC the present invention.
 XX

SQ Sequence 780 AA;

Query Match 67.6%; Score 71; DB 22; Length 780;

Best Local Similarity 81.2%; Pred. No. 0.0044;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCFCVGFSDASEPDSRH 19
 |||||
 DB 68 VGCFCVGFSDASEPDSRH 83

RESULT 6

AAW57887
 ID AAW57887 standard; Protein; 781 AA.

XX
 AC AAW57887;

DT 23-SEP-1998 (first entry)

DE Soybean raffinose synthetase.

XX
 KW Raffinose synthetase; metabolism modification; food additive;
 gastrointestinal flora; soybean.

XX
 OS Glycine max.

PN EP849359-A2.

PD 24-JUN-1998.

PF 18-DEC-1997; 97EP-0122417.

PR 18-DEC-1996; 96JP-0338673.

PA (SUMO) SUMITOMO CHEM CO LTD.

PI Oeda K, Wantanabe E;

DR WPI: 1998-324670/29.

DR N-PSDB: AAW40801.

PT New nucleic acid molecule encoding plant raffinose synthetase -
 capable of producing raffinose, used as food additives with
 beneficial effects on gastrointestinal flora

PS Claim 1; Page 31-34; 44pp; English.

XX
 CC This sequence represents the soybean raffinose synthetase of the
 CC invention. The raffinose synthetase is capable of producing raffinose by
 CC combining a D-galactosyl group through an alpha (1-6) bond with a
 CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
 CC residue in a sucrose molecule. The DNA can be used to modify metabolism
 CC of a host organism by introducing into the host organism or its cell so
 CC that the content of the raffinose family oligosaccharides in the host
 CC organism or cell is changed. Raffinose oligosaccharides are useful as
 CC food additives with beneficial effects on the gastrointestinal flora.
 XX

SQ Sequence 781 AA;

Query Match 67.6%; Score 71; DB 19; Length 781;

Best Local Similarity 81.2%; Pred. No. 0.0044;
 Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCFCVGFSDASEPDSRH 19
 |||||
 DB 68 VGCFCVGFSDASEPDSRH 83

RESULT 7

AAV30143
 ID AAV30143 standard; Protein; 781 AA.

XX
 AC AAV30143;

DT 26-OCT-1999 (first entry)

DE Amino acid sequence of a raffinose synthase protein.

XX
 KW Raffinose synthase; plant; sucrose; raffinose.

OS Glycine max.

PN JPI1215984-A.

PD 10-AUG-1999.

```
XX 12-DEC-1997; 97JP-0342899.
XX
PF 28-NOV-1997; 97JP-0329006.
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
DR WPI: 1999-51112/43.
DR N-PSDB; AAI0002.
XX
PT New raffinose synthase gene - is prepared from a plant material
XX
PS Claim 8; Page 25-27; 40pp; Japanese.
XX
CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)-D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 781 AA:

Query Match 67.6%; Score 71; DB 20; Length 781;
Best Local Similarity 81.2%; Pred. No. 0.0044;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCVFGEFADSEPDSPSRH 19
Db 68 VGCVFGEFADSEPDSPSRH 83

RESULT 8
AAB49400
ID AAB49400 standard; Protein; 781 AA.
XX
AC AAB49400;
XX
DT 07-MAR-2001 (first entry)
XX
DE Soybean raffinose synthase.
XX
KW Plant promoter; transgenic plant; desired property.
XX
OS Glycine max.
XX
PN EP1048733-A2.
XX
PD 02-NOV-2000.
XX
PF 27-APR-2000; 2000EP-0108962.
XX
PR 30-APR-1999; 99JP-0124527.
PR 01-SEP-1999; 99JP-0247211.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Ishige F, Watanabe E, Oeda K;
XX
DR WPI: 2001-104537/12.
DR N-PSDB; AAC89523.
XX
PT New soybean plant promoters useful for generating transgenic plants
PT with desired properties -
XX
PS Example 6; Page 24-27; 36pp; English.
XX
CC The present invention provides novel plant promoters which can be used in
CC the production of transgenic plants which express genes with desired
CC properties.
XX
SQ Sequence 781 AA:

Query Match 67.6%; Score 71; DB 22; Length 781;
```

```
Best Local Similarity 81.2%; Pred. No. 0.0044;
Matches 13; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 4 VGCVFGEFADSEPDSPSRH 19
Db 68 VGCVFGEFADSEPDSPSRH 83

RESULT 9
AAW57886
ID AAW57886 standard; Protein; 799 AA.
XX
AC AAW57886;
XX
DT 23-SEP-1998 (first entry)
XX
DE Broad bean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; broad bean.
XX
OS Vicia faba.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Watanabe E;
XX
DR WPI: 1998-324670/29.
DR N-PSDB; AAV40800.
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose; used as food additives with
PT beneficial effects on gastrointestinal flora.
XX
PS Claim 1; Page 26-29; 44pp; English.
XX
CC This sequence represents the broad bean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 799 AA:

Query Match 60.0%; Score 63; DB 19; Length 799;
Best Local Similarity 66.7%; Pred. No. 0.092;
Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

OY 5 GCFVGFADSEPDSPSRH 19
Db 87 GCFVGFADSEPDSPSRH 101

RESULT 10
AAAY30142
ID AAAY30142 standard; Protein; 799 AA.
XX
AC AAAY30142;
XX
DT 26-OCT-1999 (first entry)
```

```

DE  Amino acid sequence of a raffinose synthase protein.
XX
XX  Raffinose synthase; plant; broad bean; sucrose; raffinose.
XX
XX  Vicia faba.
XX  OS
XX  JP11215984-A.
XX  PD
XX  10-AUG-1999.
XX  PE
XX  12-DEC-1997; 97JP-0342899.
XX
XX  28-NOV-1997; 97JP-0329006.
XX  PR
XX  18-DEC-1996; 96JP-0338673.
XX
XX  (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX  WPI: 1999-511112/43.
XX  DR
XX  N-PSDB: AA210001.
XX
XX  New raffinose synthase gene - is prepared from a plant material
XX
XX  Claim 5; Page 19-21; 40pp; Japanese.
XX
XX  The present sequence represents a raffinose synthase protein. The
XX  CC sequence is isolated from plant material of broad beans. The
XX  CC protein forms raffinose by complexing alpha(1 to 6)- D-galactosyl
XX  CC hydroxyl group of the 6c of D-glucose residue in sucrose molecules.
XX
XX  Sequence 799 AA:
XX
XX  Query Match 60.0%; Score 63; DB 20; Length 799;
XX  Best Local Similarity 66.7%; Pred. NO. 0.092;
XX  Matches 10; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
OY 5 GCFVGFDPASEPDSRH 19
DB 87 GCFVGFNSTEPRKSH 101

RESULT 11
AAV70978
ID AAV70978 standard; Protein: 758 AA.
XX
XX  AAV70978;
XX
XX  09-AUG-2000 (first entry)
XX
XX  Soybean raffinose synthase from clone sfl1.pk125.d4.
XX  DE
XX  Soybean: raffinose synthase; raffinose saccharide;
XX  KM clone sfl1.pk125.d4; nutritional; soy protein.
XX
XX  Glycine max.
XX  OS
XX  WO200024915-A2.
XX  PN
XX  04-MAY-2000.
XX  PD
XX  22-OCT-1999; 99WO-US24923.
XX
XX  23-OCT-1998; 98US-0105451.
XX  PR
XX  (DUPO ) DU PONT DE NEMOURS & CO E. I.
XX
XX  Allen SM, Hiltz WD;
XX  PI
XX  WPI: 2000-350754/30.
XX  DR
XX  N-PSDB: AAD00335.
XX
XX  Nucleic acids and encoded proteins involved in the biosynthesis of
XX  PT raffinose, useful for producing soybean seeds with a reduced raffinose
XX  PT content and therefore improved nutritional quality -

```

```

XX
XX  Claim 2; Page 47-49; 58pp; English.
XX
XX  The present sequence is a raffinose synthase from
XX  CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
XX  CC library sfl1. Raffinose synthase is involved in the biosynthesis
XX  CC of raffinose and higher homologues in the raffinose saccharide family
XX  CC from sucrose. The present sequence is useful for reducing the raffinose
XX  CC saccharide content of soybean seeds which improves the nutritional
XX  CC quality of the soy protein products derived from them.
XX
XX  Sequence 758 AA:
XX
XX  Query Match 59.0%; Score 62; DB 21; Length 758;
XX  Best Local Similarity 58.8%; Pred. NO. 0.13;
XX  Matches 10; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
OY 3 SVGCFVGFDPASEPDSRH 19
DB 54 TTGCFLGFHATSPKSRH 70

RESULT 12
AAV32073
ID AAV32073 standard; Protein: 783 AA.
XX
XX  AAV32073;
XX
XX  17-JAN-2000 (first entry)
XX
XX  Sugarbeet raffinose synthase.
XX  DE
XX  Raffinose synthase; sugarbeet; transgenic plant.
XX
XX  Beta vulgaris.
XX
XX  Key Location/Qualifiers
XX  FH Misc-difference 227
XX  FT Misc-difference 227
XX  FT Misc-difference 348
XX  FT Misc-difference 348
XX  FT "/note=" "encoded by CCW"
XX  FT "/note=" "encoded by CCR"
XX
XX  EP953643-A2.
XX  PN
XX  03-NOV-1999.
XX  PD
XX  27-APR-1999; 99EP-0107430.
XX
XX  30-APR-1998; 98JP-0120550.
XX  PR
XX  30-APR-1998; 98JP-0120551.
XX  PR
XX  04-DEC-1998; 98JP-0345590.
XX  PR
XX  10-DEC-1998; 98JP-0351246.
XX
XX  (SUMO ) SUMITOMO CHEM CO LTD.
XX  PA
XX  Watanabe E, Oeda K;
XX  PI
XX  WPI: 1999-593144/51.
XX  DR
XX  N-PSDB: AA220208.
XX
XX  New sense and antisense genes, useful for altering the level of
XX  PT raffinose in food plants -
XX
XX  Claim 25; Page 22-24; 55pp; English.
XX
XX  This sequence represents sugarbeet raffinose synthase, a protein
XX  CC that can bind a D-galactosyl group through an alpha(1-6) bond to the
XX  CC hydroxy group attached to the carbon atom at the 6-position of the
XX  CC D-glucose residue in a sucrose molecule to form raffinose. cDNA
XX  CC (see AA220208) encoding the enzyme was isolated from sugarbeet cv.
XX  CC hanning leaf cDNA by PCR. Probes or primers generated from plant
XX  CC raffinose synthase genes (see AA220207-10) may be used to obtain
XX  CC other raffinose synthase genes by labeled detection or amplification

```

CC (claimed). These genes may be used to control the levels of
CC raffinose produced in plants. Antisense genes can be used to knock
CC out existing gene activity, and sense genes to increase the level
CC of gene activity. The resulting transgenic plants may be used as a
CC food source to alter the growing conditions for gut enterobacteria,
CC providing general health advantages.

XX Sequence 783 AA;

Query Match 58.1%; Score 61; DB 20; Length 783;

Best Local Similarity 73.3%; Pred No. 0.19;

Matches 11; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCFVGFDAEPDSRH 19

DB 60 GLFVGFDAEPKARRH 74

RESULT 13

AAY70976 AAY70976 standard; Protein: 770 AA.

XX AAY70976;

DT 09-AUG-2000 (first entry)

XX Rice raffinose synthase from clone rls24.pk0017.g10.

XX Rice: raffinose synthase; raffinose saccharide; soybean;

KW clone rls24.pk0017.g10; nutritional; soy protein.

XX Oryza sativa.

XX Key Location/Qualifiers

FT MISC-difference 100..101

FT MISC-difference 101..110

FT MISC-difference 101..110

FT /label= "Unknown

FT /note= "Encoded by NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN"

XX WO200024915-A2.

XX 04-MAY-2000.

XX 22-OCT-1999; 99WO-US24923.

XX 23-OCT-1998; 98US-0105451.

XX (DUPO) DU PONT DE NEMOURS & CO E I.

XX Allen SM, Hiltz WD;

XX WPI: 2000-350754/30.

DR N-PSDB: AAD00333.

XX Nucleic acids and encoded proteins involved in the biosynthesis of
PT raffinose, useful for producing soybean seeds with a reduced raffinose
PT content and therefore improved nutritional quality -

XX Claim 2; Page 40-42; 58pp; English.

XX The present sequence is a raffinose synthase from

CC clone rls24.pk0017.g10 isolated from a rice infected leaf cDNA

CC library rls24. Raffinose synthase is involved in the biosynthesis

CC of raffinose and higher homologues in the raffinose saccharide family

CC from sucrose. The present sequence is useful for reducing the raffinose

CC saccharide content of soybean seeds which improves the nutritional

CC quality of the soy protein products derived from them.

XX Sequence 770 AA;

Query Match 45.2%; Score 47.5; DB 21; Length 770;

Best Local Similarity 57.9%; Pred. No. 30;

Matches 11; Conservative 2; Mismatches 5; Indels 1; Gaps 1;

OY 1 PVSVCVFDPDASEPDSRH 19

DB 39 PVD-GVFICGDFAEPASRH 56

RESULT 14

AAG16865 AAG16865 standard; Protein: 175 AA.

XX AAG16865;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 17671.

XX Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

KW termination sequence.

XX Arabidopsis thaliana.

XX EPI03405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130440.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 30-APR-1999; 99US-0132048.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 07-MAY-1999; 99US-0132486.

XX 11-MAY-1999; 99US-0132863.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

XX 03-JUN-1999; 99US-0137528.

XX 04-JUN-1999; 99US-0137502.

XX 07-JUN-1999; 99US-0137724.

XX 08-JUN-1999; 99US-0138094.

XX 10-JUN-1999; 99US-0138540.

XX 10-JUN-1999; 99US-0138847.

XX 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161992.
PR 29-OCT-1999; 99US-0162142.

Query Match 44.8%; Score 47; DB 21; Length 175;
Best Local Similarity 64.3%; Pred. No. 8;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCFVGFDPASPPSR 18
DB 90 GEFLGFDPENEPISR 103

RESULT 15
AAG16864
ID AAG16864 standard; Protein; 253 AA.
XX
AC AAG16864;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 17670.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825;
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0132856.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138547.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

```
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
```

```
Query Match 44.88; Score 47; DB 21; Length 253;
Best Local Similarity 64.38; Pred. No. 12;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
```

```
OY 5 GCFVGFDPASEPDR 18
|:|:|:|:|:|:|
Db 168 GEFICGFDENEPTSR 181
```

Search completed: March 27, 2003, 10:03:43
Job time : 15.439 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 5.44512 Seconds
(without alignments)
335.448 Million cell updates/sec

Title: US-09-675-208-2

Perfect score: 105

Sequence: 1 PVSVCFFVGFDPASEPDSRH 19

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : PIR_73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	49	46.7	836	2 AF3233	conserved hypotet
2	47	44.8	295	2 B66320	3-phosphoserine ph
3	47	44.8	295	2 T51362	phosphoserine phos
4	47	44.8	336	2 A97544	nifH3/smm1 family
5	47	44.8	336	2 AB2763	conserved hypotet
6	46	43.8	483	2 AH3445	GTP-binding protei
7	44	41.9	112	2 S68584	hydrophobin A - cu
8	44	41.9	301	2 S73023	hypothetical prote
9	44	41.9	428	2 A43741	terminus protein -
10	44	41.9	498	2 T48269	hypothetical prote
11	44	41.9	650	1 JCI450	fibroblast growth
12	43	41.0	313	2 T33185	hypothetical prote
13	43	41.0	441	2 F66185	hypothetical prote
14	43	41.0	447	2 T02547	hypothetical prote
15	43	41.0	464	2 S58898	hypothetical prote
16	43	41.0	464	2 T48675	kynureninase (EC 3
17	43	41.0	478	2 T15516	hypothetical prote
18	42.5	40.5	546	2 A69890	hypothetical prote
19	42	40.0	72	2 T46623	hypothetical prote
20	42	40.0	169	2 G72559	hypothetical prote
21	42	40.0	240	2 E83896	hypothetical prote
22	42	40.0	311	2 T32776	hypothetical prote
23	42	40.0	314	2 T27383	hypothetical prote
24	42	40.0	364	2 AD2302	hypothetical prote
25	42	40.0	511	2 T23459	hypothetical prote
26	42	40.0	653	2 B87698	transketolase I [l
27	41.5	39.5	219	2 T42605	envelope protein -
28	41	39.0	125	2 E81192	hypothetical prote
29	41	39.0	314	2 T25842	hypothetical prote

30	41	39.0	511	2 G87609	L-aspartate oxidase
31	41	39.0	552	2 T16345	hypothetical prote
32	41	39.0	5232	2 A45086	HC-toxin synthetase
33	40.5	38.6	498	2 A83635	conserved hypotet
34	40	38.1	80	2 G86417	unknown protein, 6
35	40	38.1	104	2 S36510	E7 protein - human
36	40	38.1	260	2 G86228	hypothetical prote
37	40	38.1	282	2 A36875	plasmalemmal volta
38	40	38.1	362	2 A84187	hypothetical prote
39	40	38.1	447	2 T50705	hypothetical prote
40	40	38.1	465	2 G02652	gamma-aminobutyrate
41	40	38.1	520	2 AD2383	kynureninase (EC 3
42	40	38.1	609	2 E82423	Na+/H+-exchanging
43	40	38.1	615	1 ABCBS	hypothetical prote
44	40	38.1	650	2 A54976	serum albumin prec
45	40	38.1	755	2 B41836	peroxisomal matrix
					amine oxidase (fla

ALIGNMENTS

RESULT 1
AF3233
conserved hypothetical protein Atu6048 [Imported] - Agrobacterium tumefaciens (strain C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AF3233
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Genthner, D.; Kutyavina, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AF3233
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-836 <KUR>
A:Cross-references: GB:AE008690; PIDN:AAI46284.1; PID:g17744066; GSPDB:GN00189
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu6048
A:Genome: plasmid
C:Superfamily: Lactococcus lactis hypothetical protein PL08712.1
Query Match 46.7%; Score 49; DB 2; Length 836;
Best Local Similarity 41.2%; Pred. No. 6.5;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;
QY 3 SVSCFFVGFDPASEPDSRH 19
DB 687 NLGATVGFDPAREPAVH 703
RESULT 2
B66320
3-phosphoserine phosphatase [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B66320
R:Rheologos, A.; Ecker, J.R.; Palm, C.D.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dekar,
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luoro, J.S.; Malt, R.; Marzla
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:1016719; PMID:11130712
A:Accession: B66320

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-295 <STO>
A:Cross-references: GB:AE005172; NID:g97955592; PIDN:AAF98410.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 44.8%; Score 47; DB 2; Length 295;
Best Local Similarity 64.3%; Pred. No. 4.8;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCFVGFDAEPDSR 18
DB 210 GEFLEGFDENEPTSR 223

RESULT 3
T51362
phosphoserine phosphatase (EC 3.1.3.3) precursor, chloroplast [validated] - Arabidopsis
N:Alternate names: 3-phosphoserine phosphatase
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
C:Accession: T51362
R:HO, C.; Noji, M.; Saito, K.
J. Biol. Chem. 274, 11007-11012, 1999
A:Title: Plasticidic pathway of serine biosynthesis. Molecular cloning and expression of
A:Reference number: Z25385
A:Accession: T51362
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-295 <HOC>
A:Cross-references: EMBL:AB018408; PIDN:BAA33806.1
C:Genetics:
A:Gene: PSP
A:Map position: 1
A:Genome: nuclear
C:Function:
A:Description: (EC 3.1.3.3) [validated, MUID:99214175]
A:Pathway: serine biosynthesis (plasticidic pathway)
C:Keywords: chloroplast; phosphoric monoester hydrolase

Query Match 44.8%; Score 47; DB 2; Length 295;
Best Local Similarity 64.3%; Pred. No. 4.8;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 5 GCFVGFDAEPDSR 18
DB 210 GEFLEGFDENEPTSR 223

RESULT 4
nif3/smm1 family protein WC0379 [imported] - Agrobacterium tumefaciens (strain C58, Cer
A:Accession: A97544
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: A97544
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollm, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: A97544
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87306.1; PID:g15156601; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_2794
A:Map position: circular chromosome
C:Superfamily: Haemophilus Influenzae conserved hypothetical protein HI0634

Query Match 44.8%; Score 47; DB 2; Length 336;
Best Local Similarity 47.1%; Pred. No. 5.5;

Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
QY 1 PVSVCVGFDAEPDS 17
DB 135 PVTWKCRIGVDDDEPT 151

RESULT 5
AB2763
conserved hypothetical protein Atu1515 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AB2763
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AB2763
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL42520.1; PID:g17739940; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu1515
A:Map position: circular chromosome
C:Superfamily: Haemophilus Influenzae conserved hypothetical protein HI0634

Query Match 44.8%; Score 47; DB 2; Length 336;
Best Local Similarity 47.1%; Pred. No. 5.5;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCVGFDAEPDS 17
DB 135 PVTWKCRIGVDDDEPT 151

RESULT 6
AH3445
GTP-binding protein [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AH3445
R:DeIvecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanov
.; Mazur, M.; Goldsman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Let
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melit
A:Reference number: AD3252; PMID:11756688
A:Accession: AH3445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-483 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52731.1; PID:g17983561; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11550
A:Map position: I
C:Superfamily: Mycobacterium leprae probable GTP-binding protein: translation elongat

Query Match 43.8%; Score 46; DB 2; Length 483;
Best Local Similarity 47.1%; Pred. No. 12;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 1 PVSVCVGFDAEPDS 17
DB 184 PAAVGAIVGDDIEDPDA 200

RESULT 7

S68584
hydrophobin A - cultivated mushroom
C:Species: Agaricus bisporus (cultivated mushroom)
C:Date: 06-Dec-1996 #sequence_revision 07-Feb-1997 #text_change 21-Jul-2000
C:Accession: S68584; S68586; S58342
R:De Groot, P.W.J.; Schaap, P.J.; Sonnenberg, A.S.M.; Visser, J.; van Gilsen, L.J.L.D
J. Mol. Biol. 257, 1008-1018, 1996
A:Title: The Agaricus bisporus hypha gene encodes a hydrophobin and specifically accumulates
A:Reference number: S68584; MUID:96192085; PMID:8632464
A:Accession: S68584
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 1-112 <DEC>
A:Cross-references: EMBL:X90818; NID:g1235753; PIDN:CAA62331.1; PID:g1235754
A:Accession: S68586
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 1-112 <DEM>
A:Cross-references: EMBL:X89242; NID:g944820; PID:g944821
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995
C:Genetics:
A:Gene: hypha
A:Introns: 18/3; 40/1; 96/3
C:Superfamily: 1G2 protein

Query Match 41.9%; Score 44; DB 2; Length 112;
Best Local Similarity 46.2%; Pred. No. 5.5;
Matches 6; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

OY 1 PVSVCVGFVGDAS 13
I::I I::I I::I
Db 60 PNLGAFLEFDCDCT 72

RESULT 8

hypothetical protein L518_F3_81 - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C:Accession: S73023
R:Smith, D.R.; Robinson, K.
submitted to the EMBL Data Library, November 1993
A:Description: Mycobacterium leprae cosmid L518.
A:Reference number: S72591
A:Accession: S73023
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <SMT>
A:Cross-references: EMBL:U00023; NID:g467194; PIDN:AAA17366.1; PID:g467209
C:Superfamily: Mycobacterium leprae hypothetical protein L518_F3_81

Query Match 41.9%; Score 44; DB 2; Length 301;
Best Local Similarity 57.1%; Pred. No. 15;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 1 PVSVCVGFVGDASE 14
I::I I::I I::I
Db 200 PARCGCFVGDFTFE 213

RESULT 9

terminals protein - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 04-Mar-2000
C:Accession: A43741
R:Baladevall, R.M.; Mahoney, P.A.; Salas, F.; Gustavson, E.; Boyer, P.D.; Chang, M.F.; F
Dev. Biol. 125, 85-95, 1988
A:Title: Transcripts of the Drosophila blastoderm-specific locus, terminus, are concentrated
A:Reference number: A43741; MUID:88055884; PMID:3334721
A:Accession: A43741
A:Status: preliminary
A:Molecule type: mRNA

A:Residues: 1-428 <BAL>
A:Cross-references: EMBL:M19140; NID:g158535; PID:g158536
C:Genetics:
A:Gene: FlyBase:term
A:Cross-references: FlyBase:Fbgn0003683
C:Superfamily: Drosophila terminus protein
C:Keywords: DNA binding

Query Match 41.9%; Score 44; DB 2; Length 428;
Best Local Similarity 50.0%; Pred. No. 22;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

OY 6 CFVGFADSEPDSSRH 19
I::I I::I I::I
Db 37 CYSIDADPEPSQH 50

RESULT 10

hypothetical protein T22P11.70 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 19-May-2000
C:Accession: T48269
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224490
A:Accession: T48269
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-498 <BEV>
A:Cross-references: EMBL:AL162971
A:Experimental source: cultivar Columbia; BAC clone T22P11
C:Genetics:
A:Map position: 5
A:Note: T22P11.70
C:Superfamily: Arabidopsis thaliana hypothetical protein F13M22.7

Query Match 41.9%; Score 44; DB 2; Length 498;
Best Local Similarity 70.0%; Pred. No. 26;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 5 GCFVGFADSE 14
I::I I::I I::I
Db 132 GCFTGFDKSD 141

RESULT 11

fibroblast growth factor receptor 4 - rat
N:Contains: protein-tyrosine kinase (EC 2.7.1.112)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 01-Dec-2000
C:Accession: JCI450; PT0191
R:Horlick, R.A.; Stack, S.L.; Cooke, G.M.
Gene 120, 291-295, 1992
A:Title: Cloning, expression and tissue distribution of the gene encoding rat fibrobl
A:Reference number: JCI450; MUID:93013049; PMID:1398143
A:Accession: JCI450
A:Molecule type: mRNA
A:Residues: 1-650 <HOR>
A:Cross-references: GB:M91599; NID:g204137; PIDN:AAA41157.1; PID:g204138
R:La1, C.; Lemke, G.
Neuron 6, 691-704, 1991
A:Title: An extended family of protein-tyrosine kinase genes differentially expressed
A:Reference number: PT0191; MUID:91222560; PMID:2023425
A:Accession: PT0191
A:Molecule type: mRNA
A:Residues: 465-518 <LA1>
A:Experimental source: sciatic nerve
C:Genetics:
A:Gene: FGFR4; tyro-9
A:Function:
A:Description: receptor mediating effects of fibroblast growth factor

A>Note: expressed in normal lung; expressed in some carcinomas
 C:Superfamily: basic fibroblast growth factor receptor 1; immunoglobulin homology; prot
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; growth factor receptor;
 F:11-72/Domain: immunoglobulin homology <IM1>
 F:110-181/Domain: immunoglobulin homology <IM2>
 F:218-238/Domain: transmembrane #status predicted <TM>
 F:239-650/Domain: intracellular #status predicted <INT>
 F:313-598/Domain: protein kinase homology <KIN>
 F:321-329/Region: protein kinase ATP-binding motif
 F:104,136,157,168/Binding site: carboxyrate (Asn) (covalent) #status predicted
 F:351,368,460/Active site: Lys, Glu, Asp #status predicted
 F:465,478/Binding site: magnesium (Asn, Asp) #status predicted
 F:491/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status predicted

Query Match 41.9%; Score 44; DB 1; Length 650;
 Best Local Similarity 37.5%; Pred. No. 34;
 Matches 9; Conservative 3; Mismatches 4; Indels 8; Gaps 1;
 QY 1 PVSVCF-----VGPDSFPD 16
 1: |||| :|||
 Db 320 PLGCGCGVCAALGMDSPD 343

RESULT 12
 T33185
 hypothetical protein F22F7.6 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
 C:Accession: T33185
 R:Miller, N.; Kramer, J.; Smith, A.
 submitted to the EMBL Data Library, May 1998
 A:Description: The sequence of C. elegans cosmid F22F7.
 A:Reference number: Z21299
 A:Accession: T33185
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-313 <ML>
 A:Cross-references: EMBL:AF067937; PIDN:AC19217.1; GSPDB:GN00023; CESP:F22F7.6
 A:Experimental source: strain Bristol NZ; clone F22F7
 C:Genetics:
 A:Gene: CESP:F22F7.6
 A:Map position: 5
 A:introns: 18/2; 82/3; 117/1; 210/2; 241/3; 287/2

Query Match 41.0%; Score 43; DB 2; Length 313;
 Best Local Similarity 50.0%; Pred. No. 23;
 Matches 7; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 PVSVCFGFDPDASE 14
 1:||||:|:|
 Db 51 PISVGIFDFHSSQ 64

RESULT 13
 F66185
 hypothetical protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 17-May-2002
 C:Accession: F66185
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 ansen, N.F.; Hughes, B.; Hultar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Lin, Y.; Lin, S.X.; Liu, Z.A.; Luoro, J.S.; Maiti, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: AB6141; MUID:21016719; PMID:11130712
 A:Accession: F66185
 A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-441 <STO>
 A:Cross-references: GB:AE005172; NID:g238583; PIDN:AA71464.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1
 C:Superfamily: Escherichia coli probable zinc proteinase yael

Query Match 41.0%; Score 43; DB 2; Length 441;
 Best Local Similarity 57.1%; Pred. No. 33;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;
 QY 4 VCGFVGFDASPPDS 17
 1:||||:|:|
 Db 139 LGGFVGFPMNDPDS 152

RESULT 14
 T02547
 hypothetical protein Atg32480 [imported] - Arabidopsis thaliana
 N:Alternate names: hypothetical protein T26B15.4
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C:Accession: T02547; F84733
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Syke
 submitted to the EMBL Data Library, July 1998
 A:Description: Arabidopsis thaliana chromosome II BAC T26B15 genomic sequence.
 A:Reference number: Z14678
 A:Accession: T02547
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-447 <ROU>
 A:Cross-references: EMBL:AC004681; NID:g3298532; PID:g3298536
 A:Experimental source: cultivar Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, R.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: AB4420; MUID:20083487; PMID:10617197
 A:Accession: F84733
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-447 <STO>
 A:Cross-references: GB:AE002093; NID:g3298536; PIDN:AC25930.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: T26B15.4; Atg32480
 A:Map position: 2
 C:Superfamily: Escherichia coli probable zinc proteinase yael

Query Match 41.0%; Score 43; DB 2; Length 447;
 Best Local Similarity 57.1%; Pred. No. 34;
 Matches 8; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

QY 4 VCGFVGFDASPPDS 17
 1:||||:|:|
 Db 145 LGGFVGFPMNDPDS 158

RESULT 15
 S59898
 kynureninase (EC 3.7.1.3) - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Jul-1996 #sequence_revision 13-Mar-1997 #text_change 13-Mar-1997
 C:Accession: S59898
 R:Takouchi, F.; Tsubouchi, R.; Yoshino, M.; Shibata, Y.
 Biochim. Biophys. Acta 1252, 185-188, 1995
 A:Title: Amino acid sequence of rat liver kynureninase.
 A:Reference number: S59898; MUID:96049498; PMID:7578221
 A:Accession: S59898
 A:Status: preliminary
 A:Molecule type: protein
 A:Residues: 1-464 <TAK>
 C:Keywords: hydrolase

Query Match 41.0%; Score 43; DB 2; Length 464;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 5 GCEVCFD 11
 Db 244 GCEVCFD 250

Search completed: March 27, 2003, 10:06:50
 Job time : 8.44512 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 : Search time 3.12805 Seconds
(without alignments)
251,930 Million cell updates/sec

Title: US-09-675-208-2
Perfect score: 105
Sequence: 1 PVSVCFCVGFDPASEPDSRH 19

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*
Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44	41.9	112	1	HPPI_AGABI
2	44	41.9	428	1	TERM_DROME
3	43	41.0	464	1	KYNU_RAT
4	43	41.0	478	1	KYNU_CAEBL
5	42	40.0	364	1	Y3H1_ANASP
6	41.5	39.5	222	1	GL19_ARATH
7	41	39.0	5217	1	HHSI_COCCA
8	40	38.1	104	1	VE7_HPV32
9	40	38.1	282	1	POR1_BOVIN
10	40	38.1	282	1	POR1_RABIT
11	40	38.1	296	1	POR1_MOUSE
12	40	38.1	367	1	PGK_PAPPR
13	40	38.1	465	1	KYNU_HUMAN
14	40	38.1	615	1	ALBU_CHICK
15	40	38.1	650	1	PEX8_PICAN
16	40	38.1	755	1	AMO_KLEAE
17	40	38.1	834	1	DPOL_THECA
18	40	38.1	834	1	DPOL_THERM
19	40	38.1	893	1	DSCI_BOVIN
20	39.5	37.6	329	1	E2B1_MENYA
21	39.5	37.6	405	1	VGIG_HSV4
22	39.5	37.6	415	1	RL3_DROME
23	39	37.1	250	1	DAPB_CLOAB
24	39	37.1	269	1	THIG_CAUCR
25	39	37.1	301	1	MCAT_HUMAN
26	39	37.1	331	1	DEFA_MOUSE
27	39	37.1	348	1	YK23_CAEBL
28	39	37.1	389	1	BMRI_BACSU
29	39	37.1	481	1	KRI_HSV11
30	39	37.1	725	1	VG14_YEAST
31	39	37.1	752	1	8511_TRYCR
32	39	37.1	757	1	AMO_ECOLI
33	39	37.1	1088	1	RRPO_ROTBR

34	39	37.1	1088	1	RRPO_ROTBR	P21615 bovine rola
35	39	37.1	1088	1	RRPO_ROTBR	P17699 porcine rot
36	39	37.1	1088	1	RRPO_ROTBR	P22678 simian 11 r
37	39	37.1	1337	1	P152_YEAST	P39685 saeccharomyc
38	38.5	36.7	218	1	VG14_HSV4	P28941 equine herp
39	38.5	36.7	506	1	CE4B_RABIT	P15128 oryctolagus
40	38.5	36.7	1074	1	SM5A_MOUSE	O13591 homo sapien
41	38.5	36.7	1074	1	SM5A_MOUSE	O62217 mus musculu
42	38	36.2	111	1	VE7_HPV07	P36816 human papill
43	38	36.2	216	1	CSGD_SALTY	O54294 salmonella
44	38	36.2	301	1	MCAT_MOUSE	O94226 mus musculu
45	38	36.2	301	1	MCAT_RAT	P97521 rattus norv

ALIGNMENTS

RESULT 1
ID HPPI_AGABI STANDARD: PRT: 112 AA.
AC P49072:
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hydrophobin 1 precursor (Hydrophobin A).
GN HYPA OR ABH1.
OS Agaricus bisporus (Common mushroom).
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Agaricaceae; Agaricus.
OX NCBI_TaxID=5341;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Horst H39;
RX MEDLINE=96192085; PubMed=8632464;
RA de Groot P.W.J., Schaap P.J., Visser J., van Griensven L.J.L.D.:
RT "The Agaricus bisporus hypha gene encodes a hydrophobin and
RT specifically accumulates in peel tissue of mushroom caps during fruit
RT body development."
RL J. Mol. Biol. 257:1008-1018(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Horst U3;
RX MEDLINE=96254018; PubMed=8704971;
RA Luogone L.G., Bosscher J.S., Scholtmeyer K., de Vries O.M.H.,
RA Wessels J.G.H.:
RT "An abundant hydrophobin (ABH1) forms hydrophobic rodlet layers in
RT Agaricus bisporus fruiting bodies."
RL Microbiology 142:1321-1329(1996).
CC - FUNCTION: FORMS HYDROPHOBIC RODLET LAYERS.
CC - SUBCELLULAR LOCATION: Secreted.
CC - DEVELOPMENTAL STAGE: EXPRESSED IN FRUITING BODIES.
CC - SIMILARITY: BELONGS TO THE FUNGAL HYDROPHOBIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X89242; CAA61530.1; -
CC EMBL: X90818; CAA62331.1; -
CC EMBL: X92861; CAA63447.1; -
CC InterPro: IPR001338; Hydrophobin.
CC Pfam: PF01185; Hydrophobin.
CC SMART: SM00075; HYDRO. 1.
CC PROSITE: PS00956; HYDROPHOBIN: 1.
KW Fruiting body; Cell wall; Glycoprotein; Signal.
FT SIGNAL 1 19
FT CHAIN 20 112 HYDROPHOBIN 1.
FT CONFLICT 40 41 CD -> Y (IN REF. 2).
FT SEQUENCE 112 AA; 11190 MW; F7647C12826FP6E1 CRC64;

CC HYDROXYANTHRANILIC ACIDS (3-OHAA), RESPECTIVELY. HAS A PREFERENCE
CC FOR THE L-3-HYDROXY FORM. OPTIMUM ACTIVITY IS AROUND PH 9.0 FOR L-
CC KYN AND AROUND 8.5 FOR L-3OHKYN. ALSO HAS CYSTEINE-CONJUGATE-
CC BETA-LYASE ACTIVITY.
CC -1- CATALYTIC ACTIVITY: L-kynurenine + H(2)O = anthranilate + L-
CC alanine.
CC -1- CATALYTIC ACTIVITY: L-3-hydroxykynurenine + H(2)O = 3-
CC hydroxyanthranilate + L-alanine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- ENZYME REGULATION: INHIBITED BY O-METHYLBENZOTYALANINE (OMBA).
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF NAD COFACTORS FROM
CC TRYPTOPHAN THROUGH THE KYNURENINE PATHWAY.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN LIVER AND KIDNEY. ALSO
CC DETECTED IN HEART, RETINA, OVARY, LUNG, TESTIS AND BRAIN.
CC -1- INDUCTION: INHIBITED BY THIOL REAGENTS AND HEAVY METAL IONS.
CC -1- SIMILARITY: BELONGS TO THE KYNURENINASE FAMILY. SLIGHTLY RELATED
CC TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U06168; AAC53206.1; -.
CC Hydroxylase; Pyridoxal phosphate; Acetylation.
CC MOD_RES 1 ACETYLATION.
CC BINDING 276 276 PYRIDOXAL PHOSPHATE.
CC FT CONFLICT 18 18 A -> T (IN REF. 1).
CC FT CONFLICT 26 26 D -> N (IN REF. 1).
CC FT CONFLICT 118 118 T -> S (IN REF. 1).
CC SO SEQUENCE 464 AA; 52453 MW; FFI0G35E3202ECC6C CRC64;
CC
CC Query Match 41.0%; Score 43; DB 1; Length 464;
CC Best Local Similarity 100.0%; Pred. No. 14;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC Oy 5 GCFVGF 11
CC 1111111
CC Db 244 GCFVGF 250
CC
CC RESULT 4
CC KYNU_CAEL STANDARD; PRT; 478 AA.
CC ID KYNU_CAEL
CC AC Q18026;
CC DT 15-DEC-1998 (Rel. 37, Created)
CC DT 15-DEC-1998 (Rel. 37, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Probable kynureninase (EC 3.7.1.3) (L-kynurenine hydrolase).
CC GN C15H9.7.
CC OS Caenorhabditis elegans.
CC CC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
CC CC Rhabditidae; Peloderae; Caenorhabditis.
CC CC NCBI_TaxID=6239;
CC
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-Bristol N2;
CC RA Bentley D.;
CC RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN-Bristol N2;
CC RA Kohara Y., Shin'ichi T., Suzuki Y., Sugano S., Potdevin M.,
CC RA Thierry-Mieg Y., Thierry-Mieg D., Thierry-Mieg J.;
CC RA Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC CC -1- CATALYTIC ACTIVITY: L-kynurenine + H(2)O = anthranilate + L-
CC alanine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE KYNURENINASE FAMILY. SLIGHTLY RELATED
CC TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U56965; AAB52668.1; -.
CC DR EMBL: AF303267; AAC50225.1; -.
CC DR Wormpep: C15H9.7; CE06835.
CC KW Hypothetical protein; Hydroxylase; Pyridoxal phosphate.
CC FT BINDING 289 289 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC SO SEQUENCE 478 AA; 54049 MW; E889450929EC94BD CRC64;
CC
CC Query Match 41.0%; Score 43; DB 1; Length 478;
CC Best Local Similarity 100.0%; Pred. No. 15;
CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC
CC Oy 5 GCFVGF 11
CC 1111111
CC Db 257 GCFVGF 263
CC
CC RESULT 5
CC Y3H1_ANASP STANDARD; PRT; 364 AA.
CC ID Y3H1_ANASP
CC AC Q8Y064;
CC DT 15-JUN-2002 (Rel. 41, Created)
CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
CC DE Hypothetical zinc metalloprotease A113971 (EC 3.4.24.-).
CC GN A113971.
CC OS Anabaena sp. (strain PCC 7120).
CC CC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
CC CC NCBI_TaxID=103690;
CC
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=21595285; PubMed=11759840;
CC RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
CC RA Matanabe A., Iriguchi M., Ishikawa A., Kawashima K., Kimura T.,
CC RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
CC RA Nakazaki N., Shimo S., Sugimoto M., Takazawa M., Yamada M.,
CC RA Yasuda M., Tabata S.;
CC RT "Complete genomic sequence of the filamentous nitrogen-fixing
CC cyanobacterium Anabaena sp. strain PCC 7120.";
CC RL DNA Res. 8:205-213(2001).
CC CC -1- COFACTOR: zinc (Probable).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
CC (By similarity).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M50B.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP003594; BAB75670.1; -.
CC DR MEROPS: M50.004; -.
CC DR InterPro: IPR001478; PDZ.
CC DR InterPro: IPR004387; Zn_Mprotease.
CC DR InterPro: IPR000130; Zn_MTPpeptidase.
CC DR Pfam: PF00595; PDZ_1.
CC SMART: SM00228; PDZ; 1.
CC DR TIGRFAMs: TIGR00054; mem_zinc_metalprot; 1.

DR PROSITE: PS50106; PDZ; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KM Hypothetical protein: Hydrolase: Metalloprotease; Zinc; Transmembrane;
 KM Inner membrane: Complete proteome.
 FT METAL 17 17 ZINC (CATALYTIC) (POTENTIAL).
 FT ACT_SITE 18 18 POTENTIAL.
 FT METAL 21 21 ZINC (CATALYTIC) (POTENTIAL).
 FT TRANSMEM 92 114 POTENTIAL.
 FT TRANSMEM 281 303 POTENTIAL.
 FT TRANSMEM 329 346 POTENTIAL.
 FT DOMAIN 103 188 PDZ.
 SQ SEQUENCE 364 AA; 38613 MW; 54F6AE818AEFBEA CRC64;
 Query Match 40.0%; Score 42; DB 1; Length 364;
 Best Local Similarity 57.1%; Pred. No. 16;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;
 QY 4 VCGFVGFDASEPDS 17
 DB 60 LGFVGFPDDPDS 73
 RESULT 6
 GLI9_ARATH STANDARD; PRT; 222 AA.
 ID GLI9_ARATH
 AC Q9FMB0;
 DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Putative germin-like protein subfamily 1 member 9 precursor.
 GN At5G38910 OR K156.14 OR K156.90.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliopsida; Eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 OX
 RN
 SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuno A., Muraki A., Nakayama S.,
 RA Nakazaki N., Naito K., Okumura S., Shino S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spielgel L., Gao J., O'Shaughnessy A., Preston R.,
 RA Habermann K., Murray J., Johnson D., Rohlfing T., Nelson J.,
 RA Stoneking T., Pepin K., Spielth J., Sekhon M., Armstrong J., Becker M.,
 RA Belter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozeresky P., Riley A., Strommatt C.,
 RA Wagner-McPherson C., Wollam A., Yeakum M., Bell M., Dedhia N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Maria M.,
 RA Marienissen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wandut R., Duesterhoeft A., Stiekema W., Pohl T.,
 RA Lanthan K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Ramperger U., Medler H., Balke K., Medler E., Peters S.,
 RA van Staveren M., Dirkse W., Moolman P., Klein lanthorst R.,
 RA Welzenegger T., Bothe G., Rose M., Hauf J., Benneiser S., Hempel S.,
 RA Feldpausch M., Lamberth S., Villarroel R., Gietlen J., Aordles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K., Rued S., Schoof H.,
 RA Scheller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana".
 RL Nature 408:823-826(2000).
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=98290546; PubMed=9628582;
 RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,
 RA Tabata S.;

RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
 RT Sequence features of the regions of 1,456,315 bp covered by nineteen
 RT physically assigned P1 and TAC clones.";
 RL DNA Res. 5:41-54(1998).
 CC
 CC -1- FUNCTION: MAY PLAY A ROLE IN PLANT DEFENSE. HAS PROBABLY NO
 CC OXALATE OXIDASE ACTIVITY EVEN IF THE ACTIVE SITE IS CONSERVED.
 CC -1- SUBUNIT: OLIGOMER (BELIEVED TO BE A PENTAMER BUT PROBABLY
 CC HEXAMER) (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Apoplast (By similarity).
 CC -1- SIMILARITY: BELONGS TO THE GERMIN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AB009048; BAB08648.1; -
 CC HSSP: P45850; 1F12.
 DR InterPro: IPR001929; Germin.
 DR Pfam: PF01072; Germin; 1.
 DR PRINTS: PR00325; GERMIN.
 DR PROSITE: PS00725; GERMIN; 1.
 KW Apoplast; Cell wall; Signal; Glycoprotein; Manganese; Metal-binding;
 KW Multigene family; Hypothetical protein.
 FT SIGNAL 1 22
 FT CHAIN 23 222
 FT FT 111 111 SUBPARTIVE GERMIN-LIKE PROTEIN
 FT METAL 113 113 MANGANESE (BY SIMILARITY).
 FT METAL 118 118 MANGANESE (BY SIMILARITY).
 FT METAL 159 159 MANGANESE (BY SIMILARITY).
 FT DISULFID 32 48 BY SIMILARITY.
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 222 AA; 23884 MW; BB925CC4A905A4E CRC64;
 Query Match 39.5%; Score 41.5; DB 1; Length 222;
 Best Local Similarity 55.6%; Pred. No. 12;
 Matches 10; Conservative 3; Mismatches 4; Indels 1; Gaps 1;
 QY 2 VSYG-CFVGFDASEPDS 18
 DB 122 VAVGTLEVGFTVSNPENR 139
 RESULT 7
 HTS1_COCOA STANDARD; PRT; 5217 AA.
 ID HTS1_COCOA
 AC 001866;
 DT 01-FEB-1994 (Rel. 28; Created)
 DT 01-FEB-1994 (Rel. 28; Last sequence update)
 DT 15-JUN-2002 (Rel. 41; Last annotation update)
 DE HC-toxin synthetase (EC 6.3.2.-) (HTS).
 GN HTS1.
 OS Cochliobolus carbonum (Bipolaris zeicola).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; Cochliobolus.
 OC NCBI_TaxID=5017;
 OX
 RN
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.
 RC STRAIN=ATCC 90305 / SB111;
 RX MEDLINE=93100328; PubMed=1281482;
 RA Scott-Craig J.S., Panaccione D.G., Pocard J.-A., Walton J.D.;
 RT "The cyclic peptide synthetase catalyzing HC-toxin production in the
 RT filamentous fungus Cochliobolus carbonum is encoded by a
 RT 15.7-kilobase open reading frame".
 RL J. Biol. Chem. 267:26044-26049(1992).
 RN
 RP FUNCTION.
 RC STRAIN=ATCC 90305 / SB111;
 RX MEDLINE=20138231; PubMed=10671527;
 RA

RA Cheng Y.-Q., Walton J.D.;
 RT "A eukaryotic alanine racemase gene involved in cyclic peptide biosynthesis";
 RL J. Biol. Chem. 275:4906-4911(2000).
 CC -1- FUNCTION: Non-ribosomal peptide synthetase, able to activate proline and Aeo (2-amino-9,10-epoxi-8-oxodecanoic acid), and epimerize L-Pro. Catalyzes the production of HC-toxin: a cyclic tetrapeptide. Activates and thioesterifies L-Pro, and epimerizes it to D-Pro; also uses D-Ala as a substrate but this is epimerized from L-Ala by TOXG.
 CC -1- CORFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES.
 CC -1- PATHWAY: NON-RIBOSOMAL BIOSYNTHESIS OF HC-TOXIN.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME FAMILY.
 CC -1- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: M98024; AAA33023.1; -.
 CC HSSP: P14687; 1AMU.
 CC DR InterPro: IPR0010873; AMP-bind.
 CC DR InterPro: IPR001242; Condensatn.
 CC DR InterPro: IPR003880; Pantine_attach.
 CC DR Pfam: PF00501; AMP-binding; 4.
 CC DR Pfam: PF00550; pp-binding; 4.
 CC DR Pfam: PF00668; Condensation; 5.
 CC DR PRINTS: PR00154; AMPBINDING.
 CC DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 4.
 CC DR PROSITE: PS00455; AMP-BINDING; 3.
 CC DR PROSITE: PS50075; ACP_DOMAIN; 4.
 CC DR Ligase: Multifunctional enzyme: Phosphopantetheine; Repeat.
 CC KW REPEAT 249 842 DOMAIN 1.
 CC FT REPEAT 1854 2452 DOMAIN 2.
 CC FT REPEAT 3006 3606 DOMAIN 3.
 CC FT REPEAT 4158 4738 DOMAIN 4.
 CC FT DOMAIN 762 840 ACYL CARRIER (ACP) 1.
 CC FT DOMAIN 2384 2450 ACYL CARRIER (ACP) 2.
 CC FT DOMAIN 3536 3604 ACYL CARRIER (ACP) 3.
 CC FT DOMAIN 4667 4736 ACYL CARRIER (ACP) 4.
 CC FT BINDING 803 803 PHOSPHOPANTETHEINE (BY SIMILARITY).
 CC FT BINDING 2414 2414 PHOSPHOPANTETHEINE (BY SIMILARITY).
 CC FT BINDING 3568 3568 PHOSPHOPANTETHEINE (BY SIMILARITY).
 CC FT BINDING 4700 4700 PHOSPHOPANTETHEINE (BY SIMILARITY).
 CC FT BINDING 5217 5217 MW: 0331D9C5400163A5 CRC64;
 CC SO SEQUENCE

Query Match 39.0%; Score 41; DB 1; Length 5217;
 Best Local Similarity 57.1%; Pred. No. 3; 6e+02;
 Matches 8; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

OY 5 GCFVGFDAEPOS 18
 DB 4200 GCFVPLDPSTPHR 4213

RESULT 8
 VE7_HPV32
 ID VE7_HPV32 STANDARD: PRT: 104 AA.
 AC P36827;
 DT 01-JUN-1994 (rel. 29, Created)
 DT 01-JUN-1994 (rel. 29, Last sequence update)
 DT 15-JUN-2002 (rel. 41, Last annotation update)
 DE E7 protein.
 GN E7.
 OS Human papillomavirus type 32.
 OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
 OC Papillomavirus.

OX NCBI_TaxID=10612;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94265501; PubMed=8205838;
 RA Delius H., Holmann B.;
 RT "Primer-directed sequencing of human papillomavirus types";
 RL Curr. Top. Microbiol. Immunol. 186:13-31(1994).
 CC -1- FUNCTION: E7 PROTEIN HAS BOTH TRANSFORMING AND TRANS-ACTIVATING ACTIVITIES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: X74475; CAA52550.1; -.
 CC PIR: S36510; S36510.
 CC DR InterPro: IPR00148; Papv_E7.
 CC DR Pfam: PF00527; E7; 1.
 CC DR Early protein; Transcription regulation; Oncogene;
 CC KW DNA-binding; Trans-acting factor.
 CC KW SITE 64 67 C-X-C MOTIF-1.
 CC FT SITE 97 100 C-X-C MOTIF-2.
 CC SO SEQUENCE 104 AA; 11591 MW; F9DFC8C0B6D804FF CRC64;
 CC

Query Match 38.1%; Score 40; DB 1; Length 104;
 Best Local Similarity 43.8%; Pred. No. 9; 8;
 Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 PYSVGFGEFDS 16
 DB 28 PVDLYEFOFDS 43

RESULT 9
 P0R1_BOVIN
 ID P0R1_BOVIN STANDARD: PRT: 282 AA.
 AC P43879;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 16-OCT-2001 (rel. 40, Last annotation update)
 DE Voltage-dependent anion-selective channel protein 1 (VDAC-1)
 DE (Plasma membrane porin) (Brain-derived voltage-dependent anion channel 1) (BRI-VDAC).
 GN VDAC1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=94119914; PubMed=7507248;
 RA Dermietzel R., Hwang T.-K., Buettner R., Hofer A., Dotzler E., Kremer M., Deutzmann R., Thümes F.P., Fishman G.I., Spray D.C., Siemen D.;
 RT "Cloning and in situ localization of a brain-derived porin that constitutes a large-conductance anion channel in astrocytic plasma membranes";
 RT Proc. Natl. Acad. Sci. U.S.A. 91:499-503(1994).
 CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW OR ZERO MEMBRANE POTENTIAL, AND A CLOSED CONFORMATION AT POTENTIALS ABOVE 30-40 mV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY WHEREAS THE CLOSED STATE IS CATION-SELECTIVE.
 CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE OF MITOCHONDRIA AND PLASMA MEMBRANE.
 CC -1- TISSUE SPECIFICITY: PREDOMINANTLY IN BRAIN ASTROCYTES.

CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X75068; CAA52962.1; -
CC InterPro: IPR001925; Porin_Euk.
CC Pfam: PF01459; Euk_porin.1.
CC PRINTS: PR00185; EUKARYTPORIN.
CC PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
CC KMW plasma; Porin; Acetylation.
CC FT INIT MET 0 0 ACETYLATION (BY SIMILARITY).
CC FT MOD_RES 1 1
CC SQ SEQUENCE 282 AA; 30694 MW; 2E6B2055FCD7849 CRC64;
QY
QY 2 VSVGCFVGFDPASEPDSR 18
QY 122 INLGCDVDFDIAGPSIR 138
Db
RESULT 10
PORL_RABIT STANDARD; PRT; 282 AA.
ID PORL_RABIT
AC 09T15;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-dependent anion-selective channel protein 1 (VDAC-1) (Outer
DE mitochondrial membrane protein porin 1).
GN VDAC1.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=New Zealand white; TISSUE=Corneal endothelium;
RA Rae J.L.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
CC SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW
CC OR ZERO MEMBRANE POTENTIAL AND A CLOSED CONFORMATION AT POTENTIALS
CC ABOVE 30-40 mV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE OF MITOCHONDRIA AND PLASMA
CC MEMBRANE (BY SIMILARITY).
CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AF209725; AAF22835.1; -
CC InterPro: IPR001925; Porin_Euk.
CC Pfam: PF01459; Euk_porin.1.
CC PRINTS: PR00185; EUKARYTPORIN.
CC PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
CC PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
DR

KW Outer membrane; Porin; Mitochondrion; Acetylation.
FT INIT_MET 0 0 BY SIMILARITY.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT BINDING 72 72 DICYLOHEXYLCARBODIIMIDE.
SQ SEQUENCE 282 AA; 30609 MW; 59F77E96F108A298 CRC64;
QY
QY 2 VSVGCFVGFDPASEPDSR 18
QY 122 INLGCDVDFDIAGPSIR 138
Db
RESULT 11
PORL_MOUSE STANDARD; PRT; 296 AA.
ID PORL_MOUSE
AC 060932;
DT 15-JUL-1999 (Rel. 38, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Voltage-dependent anion-selective channel protein 1 (VDAC-1) (mVDAC1)
DE (mVDAC5) (Outer mitochondrial membrane protein porin 1) (Plasmalemmal
DE porin).
GN VDAC1 OR VDAC5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognath; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM MT-VDAC1).
RC TISSUE=Brain;
RX MEDLINE=96301405; Pubmed=8660977;
RA Sampson M.J., Lovell R.S., Craigen W.J.;
RT "Isolation, characterization, and mapping of two mouse mitochondrial
RT voltage-dependent anion channel isoforms.";
RN Genomics 33:283-288(1996).
RL [2]
RP CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX MEDLINE=20202612; Pubmed=10716730;
RA Buettner R., Papoutsoglou G., Seemes E., Spray D.C., Dermietzel R.;
RT "Evidence for secretory pathway localization of a voltage-dependent
RT anion channel isoform.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:3201-3206(2000).
CC -1- FUNCTION: FORMS A CHANNEL THROUGH THE MITOCHONDRIAL OUTER MEMBRANE
CC AND ALSO THE PLASMA MEMBRANE. THE CHANNEL ALLOWS DIFFUSION OF
CC SMALL HYDROPHILIC MOLECULES; IT ADOPTS AN OPEN CONFORMATION AT LOW
CC OR ZERO MEMBRANE POTENTIAL AND A CLOSED CONFORMATION AT POTENTIALS
CC ABOVE 30-40 mV. THE OPEN STATE HAS A WEAK ANION SELECTIVITY
CC WHEREAS THE CLOSED STATE IS CATION-SELECTIVE.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL VDAC1 (MT-VDAC1) IN OUTER
CC MEMBRANE OF MITOCHONDRIA AND PLASMALEMMAL VDAC1 (PL-VDAC1) IN
CC PLASMA MEMBRANE.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; PL-VDAC1 (SHOWN HERE) AND MT-
CC VDAC1; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGH LEVELS OF EXPRESSION DETECTED IN HEART,
CC KIDNEY, BRAIN, AND SKELETAL MUSCLE. NOT EXPRESSED IN TESTIS.
CC -1- DOMAIN: CONSISTS MAINLY OF MEMBRANE-SPANNING SIDED BETA-SHEETS.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC MITOCHONDRIAL PORIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U30840; AAA87777.1; -
CC SWISS-2DPAGE: 060932; MOUSE.
CC MGD: MGI:106919; Vdac1.
CC InterPro: IPR001925; Porin_Euk.
DR


```

DR Pfam: PF01459; Euk_porin; 1.
DR PRINTS: PR00185; EUKARYOTIC_PORIN.
DR PROSITE: PS00558; EUKARYOTIC_PORIN; 1.
KW Outer membrane; Porin; Mitochondrion; Alternative splicing;
KW Acetylation.
FT MOD_RES 14 14 ACETYLATION (BY SIMILARITY).
FT BINDING 86 86 DICICLOHEXYLCARBODIIMIDE (BY SIMILARITY).
FT VARSPIC 1 13 MISSING (IN ISOFORM MT-VDAC1).
SO SEQUENCE 296 AA; 32351 MW; C0710C1717063B32 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 296;
Best Local Similarity 41.2%; Pred. No. 29;
Matches 7; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

Oy 2 VSGGFGVGFDASEPDSR 18
Db 136 INLCGVDFDIAGPSIR 152

RESULT 12
PGK_PAPPR STANDARD: PRT; 367 AA.
AC 000869;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Phosphoglycerate kinase (EC 2.7.2.3) (Fragment).
GN PKR.
OS Parametium primaurella.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Penicillia;
OC Parametium.
OX NCBI_TaxID=5886;
RN [1]
RP SEQUENCE FROM N.A.
RA Pearlman R.E.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate -> ADP + 3-phospho-D-glyceroyl phosphate.
CC -1- PATHWAY: Second phase of glycolysis; second step.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF001849; AAB58241.1; -.
DR HSSP: P00560; 10PG.
DR InterPro: IPR001576; PGK.
DR Pfam: PF00162; PGK; 1.
DR PROSITE: PS00111; POLYMERASE_KINASE; PARTIAL.
KW transferase; Kinase; Glycolysis.
FT NON_TER 1 1
FT NON_TER 367 367
SO SEQUENCE 367 AA; 39876 MW; DEF64500127DE460 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 367;
Best Local Similarity 42.1%; Pred. No. 36;
Matches 8; Conservative 1; Mismatches 10; Indels 0; Gaps 0;

Oy 1 PVSVCFGVGFDASEPDSR 19
Db 259 PTFVCGTGLDASSPVALH 277

RESULT 13
KYNH_HUMAN STANDARD: PRT; 465 AA.
AC Q16719;

```

```

DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-OCT-2001 (Rel. 40, Last annotation update)
DE kynureninase (EC 3.7.1.3) (L-kynurenine hydrolase).
GN KYNH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC TISSUE=Hepatosoma;
RX MEDLINE=96314506; PubMed=8706755;
RA Alberati-Giani D., Buchli R., Malherbe P., Broger C., Lang G.,
RA Koehler C., Lahm H.-W., Cesura A.M.;
RT "Isolation and expression of a cDNA clone encoding human
RT kynureninase."
RL Eur. J. Biochem. 239:460-468(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=97324088; PubMed=9180257;
RA Toma S., Nakamura M., Tone S., Okuno E., Kido R., Breton J.,
RA Avanzi N., Cozzi L., Speciale C., Mostardini M., Gatti S., Benatti L.;
RT "Cloning and recombinant expression of rat and human kynureninase."
RL FEBS Lett. 408:5-10(1997).
CC -1- FUNCTION: CATALYZES THE CLEAVAGE OF L-KYNURENINE (L-KYN) AND L-3-HYDROXYKYNURENINE (L-3OHKYN) INTO ANTHRANILIC (AA) AND 3-HYDROXYANTHRANILIC ACIDS (3-OHAA), RESPECTIVELY. HAS A PREFERENCE FOR THE L-3-HYDROXY FORM. ALSO HAS CYSTEINE-CONJUGATE-BETA-LYASE ACTIVITY (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: L-kynurenine + H(2)O -> anthranilate + L-alanine.
CC -1- CATALYTIC ACTIVITY: L-3-hydroxykynurenine + H(2)O -> 3-hydroxyanthranilate + L-alanine.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- ENZYME REGULATION: INHIBITED BY O-METHOXYBENZYLALANINE (OMBA).
CC -1- PATHWAY: INVOLVED IN THE BIOSYNTHESIS OF NAD COFACTORS FROM TRYPTOPHAN THROUGH THE KYNURENINE PATHWAY.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN ALL TISSUES TESTED (HEART, BRAIN PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS).
CC HIGHEST LEVELS FOUND IN PLACENTA, LIVER AND LUNG. EXPRESSED IN ALL BRAIN REGIONS.
CC -1- INDUCTION: INCREASED LEVELS IN SEVERAL CEREBRAL AND SYSTEMIC INFLAMMATORY CONDITIONS.
CC -1- SIMILARITY: BELONGS TO THE KYNURENINASE FAMILY. SLIGHTLY RELATED TO CLASS-V OF PYRIDOXAL-PHOSPHATE-DEPENDENT AMINOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U57721; AAC50650.1; -.
DR Genew; HGNC:6469; KYNH.
DR MIM: 605197; -.
DR MIM: 236800; -.
KW Hydrolase; Pyridoxal phosphate; Acetylation.
FT MOD_RES 1 1 ACETYLATION (BY SIMILARITY).
FT BINDING 276 276 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SO SEQUENCE 465 AA; 52351 MW; BDD136BE18C79EBB CRC64;

Query Match 38.1%; Score 40; DB 1; Length 465;
Best Local Similarity 85.7%; Pred. No. 45;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 5 GCFVGF 11

```

DB 244 GCYVGD 250

RESULT 14

ALBU_CHICK STANDARD: PRT: 615 AA.

ID ALBU_CHICK

AC P19121

DT 01-NOV-1990 (Rel. 16, Created)

DT 01-AUG-1992 (Rel. 23, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Serum albumin precursor.

GN ALB

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI_Taxid=9031;

RN [1]

RE SEQUENCE FROM N.A.

RC TISSUE=Liver;

RA Cassidy A.I., Salkid C.K., Bayersstock P., Wallace J.C.;

RL Submitted (JUL-1991) to the EMBL/Genbank/DBJ databases.

RN [2]

RP SEQUENCE OF 1-28 FROM N.A.

RX MEDLINE=83161037; PubMed=6187737;

RA Hache R.J.G., Wiskocil R., Vasa M., Roy R.N., Lau P.C.K., Deeley R.G.;

RT "The 5' noncoding and flanking regions of the avian very low density

RT apolipoprotein II and serum albumin genes. Homologies with the egg

RT white protein genes.";

RL J. Biol. Chem. 258:4556-4564(1983).

RN [3]

RP SEQUENCE OF 19-30.

RX MEDLINE=78019943; PubMed=911327;

RA Rosen A.M., Geller D.M.;

RT "Chicken microsomal albumin: amino terminal sequence of chicken

RT proalbumin.";

RL Biochem. Biophys. Res. Commun. 78:1060-1066(1977).

CC -I- BINDING CAPACITY FOR WATER, CA⁺⁺, NA⁺, K⁺, FATTY ACIDS, HORMONES,

CC BILIRUBIN AND DRUGS. ITS MAIN FUNCTION IS THE REGULATION OF THE

CC COLLOIDAL OSMOTIC PRESSURE OF BLOOD.

CC -I- SUBCELLULAR LOCATION: Secreted.

CC -I- TISSUE SPECIFICITY: PLASMA.

CC -I- SIMILARITY: BELONGS TO THE ALB/AFP/VDB FAMILY.

CC -I- SIMILARITY: CONTAINS 3 ALBUMIN DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: X60688; CAA43098.1; -

CC EMBL: V00381; CAA23680.1; -

CC PIR: S15571; ABCS.

DR HSSP: P02768; 1E7B.

DR InterPro: IPR000264; Serum_albumin.

DR Pfam: PF00273; transport_prot. 3.

DR PRINTS: PRO0802; SERUMALBUMIN.

DR PRODOM: PD002486; Serum_albumin; 1.

DR SMART: SM00103; ALBUMIN; 3.

DR PROSITE: PS00212; ALBUMIN; 3.

KW Plasma; Metal-binding; Lipid-binding; Albumin; Repeat; Signal;

KW Copper.

FT SIGNAL 1 18

FT PROPEP 19 23

FT CHAIN 24 615 SERUM ALBUMIN.

FT DOMAIN 24 209 ALBUMIN 1.

FT DOMAIN 216 401 ALBUMIN 2.

FT DOMAIN 408 599 ALBUMIN 3.

FT METAL 30 30 COPPER (BY SIMILARITY).

FT DISULFID 80 89 BY SIMILARITY.

FT DISULFID 102 118 BY SIMILARITY.

FT DISULFID 117 128 BY SIMILARITY.

FT DISULFID 152 197 BY SIMILARITY.

FT DISULFID 196 205 BY SIMILARITY.

FT DISULFID 228 274 BY SIMILARITY.

FT DISULFID 273 281 BY SIMILARITY.

FT DISULFID 293 307 BY SIMILARITY.

FT DISULFID 306 317 BY SIMILARITY.

FT DISULFID 344 389 BY SIMILARITY.

FT DISULFID 388 397 BY SIMILARITY.

FT DISULFID 420 466 BY SIMILARITY.

FT DISULFID 465 476 BY SIMILARITY.

FT DISULFID 489 505 BY SIMILARITY.

FT DISULFID 504 515 BY SIMILARITY.

FT DISULFID 542 587 BY SIMILARITY.

FT DISULFID 586 595 BY SIMILARITY.

FT CARBOHYD 500 500 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CONFLICT 24 24 F -> M (IN REF. 3).

SQ SEQUENCE 615 AA; 69918 MW; E59E4BACAC066C6 CRC64;

Query Match 38.1%; Score 40; DB 1; Length 615;

Best Local Similarity 54.5%; Pred No. 60;

Matches 6; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 6 CFVGFDPASEPD 16

DB 128 CFLSFVQSQPD 138

RESULT 15

PEX8_PICAN STANDARD: PRT: 650 AA.

ID PEX8_PICAN

AC 000925;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Peroxisomal matrix protein PER1 precursor (Peroxin-8).

GN PEX8 OR PER1.

OS Pichia angusta (Yeast) (Hansenula polymorpha).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Pichia.

OX NCBI_Taxid=4905;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=CBS 4732;

RX MEDLINE=95050945; PubMed=7962056;

RA Waterham H.R., Titorenko V.I., Halma P., Gregg J.M., Harder W.,

RA Veenhuis M.;

RT "The Hansenula polymorpha PER1 gene is essential for peroxisome

RT biogenesis and encodes a peroxisomal matrix protein with both

RT carboxy- and amino-terminal targeting signals.";

RL J. Cell Biol. 127:737-749(1994).

CC -I- FUNCTION: ESSENTIAL FOR PEROXISOME BIOGENESIS. MAY PLAY A ROLE IN

CC TRIGGERING THE PROTEIN IMPORT COMPETENCE OF INDIVIDUAL

CC PEROXISOMES. IT MAY INTERACT WITH PEX8 (PEX10).

CC -I- SUBCELLULAR LOCATION: Peroxisomal; matrix.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL: Z30206; CAA82928.1; -

DR PROSITE: PS00342; MICROBODIES_CTER. 1.

KW Peroxisome; Transit peptide.

FT TRANSIT 1 650 MICROBODY.

FT CHAIN 2 650 PEROXISOMAL MATRIX PROTEIN PER1.

FT SITE 648 650 MICROBODY TARGETING SIGNAL (POTENTIAL).

SQ SEQUENCE 650 AA; 74122 MW; A9AC534204F50CTD CRC64;

Query Match 38.1%; Score 40; DB 1; Length 650;
 Best Local Similarity 36.8%; Pred. No. 64;
 Matches 7; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

Oy 1 PVSVCCEVGFDASEPPSRH 19
 ||:||||:|:
 Db 145 PVAVGCLLSVDSRNDYDRY 163

Search completed: March 27, 2003, 10:04:17
 Job time : 5.12805 secs

GenCore version 5.1.4-p5-4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 10.6585 seconds
(without alignments)
367.301 Million cell updates/sec

Title: US-09-675-208-2
Perfect score: 105
Sequence: 1 PVSVCFCVGFDAEPDSRH 19

Scoring table: BLAST62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-misc:*
8: SP-organelle:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*
15: SP-virus:*
16: SP-bacteriap:*
17: SP-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	105	100.0	784	10 Q92762	Q92762 cucumis sat
2	54	51.4	798	10 Q8VW6	Q8VW6 pisum sativ
3	50	47.6	779	10 Q9M4M7	Q9M4M7 persia amer
4	49	46.7	836	16 Q8U687	Q8U687 agrobacteri
5	47	44.8	295	10 Q9F285	Q9F285 arabidopsis
6	47	44.8	295	10 Q82796	Q82796 arabidopsis
7	47	44.8	336	16 Q8U684	Q8U684 agrobacteri
8	46.5	44.3	783	10 Q8FND9	Q8FND9 arabidopsis
9	46	43.8	483	16 Q8YFH2	Q8YFH2 bruceella me
10	45	42.9	199	11 Q8VEA8	Q8VEA8 mus musculu
11	45	42.9	757	13 Q9YHD2	Q9YHD2 gallus gall
12	44	41.9	202	11 Q9QVW8	Q9QVW8 gallus sp.
13	44	41.9	301	16 Q49942	Q49942 mycobacteri
14	44	41.9	421	5 Q9VWQ2	Q9VWQ2 drosophila
15	44	41.9	498	10 Q9LZ54	Q9LZ54 arabidopsis
16	44	41.9	508	10 Q9C5M0	Q9C5M0 arabidopsis

17	44	41.9	650	11 Q63709	Q63709 rattus ratt
18	44	41.9	742	10 Q9AV10	Q9AV10 oryza sativ
19	44	41.9	816	10 Q942P6	Q942P6 oryza sativ
20	43	41.0	121	17 Q971R6	Q971R6 sulfobobus
21	43	41.0	167	9 Q9KCH9	Q9KCH9 streptococc
22	43	41.0	180	5 Q9U249	Q9U249 caenorhabdi
23	43	41.0	236	10 Q9XEQ8	Q9XEQ8 sorghum bic
24	43	41.0	242	17 Q978G2	Q978G2 thermoplasm
25	43	41.0	299	10 Q94AV8	Q94AV8 arabidopsis
26	43	41.0	313	5 Q9CZP1	Q9CZP1 caenorhabdi
27	43	41.0	323	11 Q8R0T7	Q8R0T7 mus musculu
28	43	41.0	359	16 Q92P85	Q92P85 rhizobium m
29	43	41.0	441	10 Q23053	Q23053 arabidopsis
30	43	41.0	447	10 Q80885	Q80885 arabidopsis
31	43	41.0	464	11 Q9CXF0	Q9CXF0 mus musculu
32	42.5	40.5	546	16 Q31813	Q31813 bacillus su
33	42	40.0	169	17 Q9YB34	Q9YB34 aeropyrum p
34	42	40.0	240	16 Q9KBE5	Q9KBE5 bacillus ha
35	42	40.0	257	2 Q51720	Q51720 propionibac
36	42	40.0	298	5 Q9XW86	Q9XW86 caenorhabdi
37	42	40.0	311	5 Q44789	Q44789 caenorhabdi
38	42	40.0	332	16 Q98M15	Q98M15 rhizobium l
39	42	40.0	376	10 Q9FVH5	Q9FVH5 prunus arme
40	42	40.0	393	2 Q9LCH6	Q9LCH6 streptococ
41	42	40.0	419	2 Q08410	Q08410 sheanella
42	42	40.0	427	17 Q8U1G8	Q8U1G8 pyrococcus
43	42	40.0	439	2 Q52785	Q52785 actinobact
44	42	40.0	511	5 Q9XUS4	Q9XUS4 caenorhabdi
45	42	40.0	648	10 Q9M824	Q9M824 arabidopsis

ALIGNMENTS

RESULT 1					
ID Q92762	PRELIMINARY:	PRT:	784 AA.		
AC Q92762:					
DT 01-MAY-1999 (TREMBLrel. 10, Created)					
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)					
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)					
DE Raffinose synthase (EC 2.4.1.82).					
GN RFS.					
OS Cucumis sativus (Cucurbitaceae).					
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;					
OC Cucurbitales; Cucurbitaceae; Cucumis.					
OX NCBI_TaxID=3659;					
RN [1]					
RP SEQUENCE FROM N.A.					
RC TISSUE=LEAF:					
RA Ohsumi C., Nozaki J., Kida T.;					
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.					
DR EMBL: AF073744; AAD02832.1; -					
KW Glycosyltransferase; Transferase.					
SQ SEQUENCE 784 AA; 86920 MW; 3B06A491F0908933 CRC64;					
Query Match	100.0%;	Score 105;	DB 10;	Length 784;	
Best Local Similarity	100.0%;	Pred. No. 7.7e-05;			
Matches 19;	Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1 PVSVCFCVGFDAEPDSRH 19					
DB 61 PVSVCFCVGFDAEPDSRH 79					
RESULT 2					
ID Q8VW6	PRELIMINARY:	PRT:	798 AA.		
AC Q8VW6:					
DT 01-MAR-2002 (TREMBLrel. 20, Created)					
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)					
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)					

```

DE  Raffinose synthase (EC 2.4.1.82).
GN  RFS.
OS  Pisum sativum (Garden pea).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
OX  NCBI_TaxID=3888;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CV. WUNDER VON KETVEDON; TISSUE=SEED;
RC  Peterhauer T., Mech L., Mucha J., Richter A.;
RT  "Molecular characterization of raffinose synthase from pea (Pisum
RT  sativum L.) seeds.";
RL  Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AJ426475; CAD20127.2; -.
KW  Transferrase; Glycosyltransferase.
SQ  SEQUENCE 798 AA; 88717 MW; 8D3F3ED5BF617B7 CRC64;

Query Match          51.4%; Score 54; DB 10; Length 798;
Best Local Similarity 60.0%; Pred. No. 2.2;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  5 GCFVGFDPASEPDSRH 19
Db  84 GCFVGFDPTEAKSHH 98

RESULT 3
O9M4M7  PRELIMINARY; PRT; 779 AA.
AC  O9M4M7;
DT  01-OCT-2000 (TREMBlrel. 15, Created)
DT  01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DR  01-DEC-2001 (TREMBlrel. 15, Last annotation update)
DE  Putative seed imbibition protein.
GN  SIP.
OS  Persea americana (Avocado).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX  NCBI_TaxID=3435;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=CV. HASS; TISSUE=MESOCARP;
RC  Zamorano J.P., Evans A.D., Dopico B., Lowe A.L., Wilson I.D.,
RA  Merodio C., Grierson D.;
RT  "Isolation and characterization of cDNAs for mRNAs regulated during
RT  cold storage of avocado (Persea americana Mill.) fruit.";
RL  Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL; AJ133148; CAB7245.1; -.
SQ  SEQUENCE 779 AA; 85368 MW; C3A8BA3160316785 CRC64;

Query Match          47.6%; Score 50; DB 10; Length 779;
Best Local Similarity 60.0%; Pred. No. 9.7;
Matches 9; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  5 GCFVGFDPASEPDSRH 19
Db  42 GCFVGFDPASESESIH 56

RESULT 4
O8U687  PRELIMINARY; PRT; 836 AA.
AC  O8U687;
DT  01-JUN-2002 (TREMBlrel. 21, Created)
DT  01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DE  Hypothetical protein Atu6048.
GN  Atu6048 OR AGR_PTI_100.
OS  Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC  Plasmid pTiC58.
OC  Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC  Rhizobiaceae; Rhizobium.

```

```

OX  NCBI_TaxID=176299;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21608550; PubMed=11743193;
RA  Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA  Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA  Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boeve D. Sr.,
RA  Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA  Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA  Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA  Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA  Gordon Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA  Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA  Nester E.W.;
RT  "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT  C58.";
RL  Science 294:2317-2323(2001).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=21608551; PubMed=11743194;
RA  Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA  Gurolo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA  Houniel K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F.,
RA  Flanagan C., Crowell C., Gursun J., Lomo C., Seear C., Markez B.,
RA  Cielo C., Slater S.;
RT  "Genome sequence of the plant pathogen and biotechnology agent
RT  Agrobacterium tumefaciens C58.";
RL  Science 294:2323-2328(2001).
DR  EMBL; AE009423; AAL46284.1; -.
DR  EMBL; AE007931; AAK91006.1; -.
KW  Hypothetical protein; Plasmid; Complete proteome.
SQ  SEQUENCE 836 AA; 91964 MW; 3C61FD8318D3DF1 CRC64;

Query Match          46.7%; Score 49; DB 16; Length 836;
Best Local Similarity 41.2%; Pred. No. 15;
Matches 7; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY  3 SVGCFVGFDPASEPDSRH 19
Db  687 NLGAVLGFDAERPEAVH 703

RESULT 5
O9FZ85  PRELIMINARY; PRT; 295 AA.
AC  O9FZ85;
DT  01-MAR-2001 (TREMBlrel. 16, Created)
DT  01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT  01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE  3-phosphoserine phosphatase (Putative 3-phosphoserine
DE  phosphatase).
GN  F26116.2 OR ATG18640.
OS  Arabidopsis thaliana (Mouse-ear cress).
OC  Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX  NCBI_TaxID=3702;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA  Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA  Buehler E., Chao Q., Chin C., Chiu J., Choi E., Gonzalez A.,
RA  Howng B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA  Lenz C., Liu A., Liu S., Mukharsy N., Pham P., Sakano H., Shinn P.,
RA  Toriumi M., Vaysberg M., Yu G., Becker J., Theologis A., Davis R.W.;
RL  Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RN  [2]
RP  SEQUENCE FROM N.A.
RA  Yamada K., Banh J., Chang C.H., Chang E., Dale J.M., Goldsmith A.D.,
RA  Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M., Wu H.C.,
RA  Yamamura Y., Yu G., Yu S., Bowser L., Carinici P., Chen H., Cheuk R.,
RA  Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,

```

```

RA Kawai J., Kim C., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Shinozaki K., Davis R.W., Ecker J.R.,
RA Theologis A.;
RT "Full length cDNA of gene At1g18640 (GI:15221897)";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC026238; AAF98410.1; -
DR EMBL; AY065351; AAL38792.1; -
DR InterPro; IPR001454; Hlgname/hydrlase.
DR InterPro; IPR004469; SerB.
DR Pfam; PF00702; Hydrolase; 1.
DR TIGRfams; TIGR00338; serB; 1.
SQ SEQUENCE 295 AA: 32318 MW: F14C95E636E7745E CRC64;

Query Match 44.8%; Score 47; DB 10; Length 295;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCFVGFDSPPSR 18
DB 210 GFLGFDENEPTSR 223

RESULT 6
ID 082796 PRELIMINARY; PRT; 295 AA.
AC 082796;
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE 3-PHOSPHOSERIN phosphatase.
GN PSP.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Ho C., Noji M., Saito K.;
RT "Molecular cloning and characterization of PSP.";
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB018408; BAA33807.1; -
DR EMBL; AB018408; BAA33806.1; -
DR InterPro; IPR001454; Hlgname/hydrlase.
DR InterPro; IPR004469; SerB.
DR Pfam; PF00702; Hydrolase; 1.
DR TIGRfams; TIGR00338; serB; 1.
SQ SEQUENCE 295 AA: 32302 MW: F14C95E636E6754E CRC64;

Query Match 44.8%; Score 47; DB 10; Length 295;
Best Local Similarity 64.3%; Pred. No. 11;
Matches 9; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 5 GCFVGFDSPPSR 18
DB 210 GFLGFDENEPTSR 223

RESULT 7
ID 080F84 PRELIMINARY; PRT; 336 AA.
AC 080F84;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical protein At1u1515.
GN ATU1515 OR AGR_C_2794.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RX MEDLINE-21608550; PubMed-11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eissen J.A., Karp P.D., Boyce D. St.,
RA Chapman P., Clendenning J., Deatherage G., Gilliet W., Grant C.,
RA Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
C58.";
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-21608551; PubMed-11743194;
RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houniel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Mollam C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
RA Planagan C., Crowell C., Gurson J., Lomo C., Sear C., Strub G.,
RA Cleio C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58.";
RL Science 294:2323-2328(2001).
DR EMBL; AE009110; AAI42520.1; -
DR EMBL; AE008075; AAK87306.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 336 AA: 36937 MW: C06405E207BF643A CRC64;

Query Match 44.8%; Score 47; DB 16; Length 336;
Best Local Similarity 47.1%; Pred. No. 13;
Matches 8; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

OY 1 PVSVCGEVGFDSPPSR 17
DB 135 PVTVCRCIGVDQDEPET 151

RESULT 8
ID 09FEND9 PRELIMINARY; PRT; 783 AA.
AC 09FEND9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Raffinose synthase protein.
GN MPO12.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-COLUMBIA;
RX MEDLINE-98069011; PubMed-9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned P1 clones.";
RL DNA Res. 4:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;

```

RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bosser L., Jones T., Banh J., Carrincci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shin P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.,
 RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB006702: BAB1595.1; -;
 DR EMBL: AY062781: AAL3285.1; -;
 DR EMBL: AY081645: AAM10207.1; -;
 SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

Query Match
 Best Local Similarity 44.3%; Score 46.5; DB 10; Length 783;
 Matches 10; Conservative 2; Mismatches 6; Indels 1; Gaps 1;

OY 2 VSVCGFVGFDA-SEPD5RH 19
 DB 62 VSAGSEIFGNLGEPRSH 80

RESULT 9
 O8YFH2 PRELIMINARY; PRT; 483 AA.
 AC O8YFH2;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 GN BME11550.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA Ivanova N., Anderson T., Bhattacharyya A., Lykidis A., Reznik G.,
 RA Jablonksi L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J.,
 RA Haselkorn R., Kyriades N., Overbeek R.,
 RT "The genome sequence of the facultative intracellular pathogen
 RT Brucella melitensis.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 DR EMBL: AE009590; AAL52731.1; -;
 DR InterPro: IPR005289; GTP-binding_dom.
 DR InterPro: IPR002917; MMR_HSR1.
 DR InterPro: IPR005225; SMALL_GTP.
 DR Pfam: PF01926; MMR_HSR1.1.
 DR TIGRfams: TIGR00650; MG442; 2.
 DR TIGRfams: TIGR00231; small_GTP; 2.
 KW Complete proteome.
 SQ SEQUENCE 483 AA; 53311 MW; 1624111DB29AA266 CRC64;

Query Match
 Best Local Similarity 43.8%; Score 46; DB 16; Length 483;
 Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 PVSVCGFVGFDA5EPDS 17
 DB 184 PAAVGAIVGDIEDPDA 200

RESULT 10
 O8VEA8 PRELIMINARY; PRT; 199 AA.
 AC O8VEA8;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)

DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Hypothetical 22.5 kDa protein.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC019395; AAH19395.1; -;
 DR InterPro: IPR003579; GTPase_Rab.
 DR InterPro: IPR001806; Ras_transmng.
 DR InterPro: IPR000834; Zn_cardoepet.
 DR Pfam: PF00071; ras; 1.
 DR PRINTS: PR00449; RASTRNSFRMNG.
 DR SMART: SM00175; RAB; 1.
 DR PROSITE: PS00133; CARBOXYPEPT_2N_2; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 199 AA; 22502 MW; 33353842079CF6FC CRC64;

Query Match
 Best Local Similarity 42.9%; Score 45; DB 11; Length 199;
 Matches 7; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

OY 3 SVSCGFVGFDA5EPDS 17
 DB 81 SDCCIAAFDVTDPES 95

RESULT 11
 O9YHD2 PRELIMINARY; PRT; 757 AA.
 AC O9YHD2;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Nuclear calmodulin-binding protein (Fragment).
 GN URP.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RA Lodge A.P., Walsh A., McNamee C.J., Moss D.J.;
 RT "Identification of cHURP, a Nuclear Calmodulin-Binding Protein Related
 RT to HnRNP-U.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF098788; AAC69888.1; -;
 DR InterPro: IPR003878; SPRY_domain.
 DR InterPro: IPR003877; SPRY_receptot.
 DR Pfam: PF00622; SPRY; 1.
 DR SMART: SM00449; SPRY; 1.
 FT NON_TER
 SQ SEQUENCE 757 AA; 84179 MW; 7FA04ED9E176D1D CRC64;

Query Match
 Best Local Similarity 42.9%; Score 45; DB 13; Length 757;
 Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

OY 4 VSCGFVGFDA5EPD 16
 DB 342 ICGFADFDEASEE 354

RESULT 12
 O9QVW8 PRELIMINARY; PRT; 202 AA.
 AC O9QVW8;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)

Best Local Similarity 50.0%; Pred. No. 51;
Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

QY 6 CFEVGFDPDSRH 19
|: |||||:
Db 36 CYSIDADEPDSOH 49

RESULT 15

09LZ54
ID 09LZ54 PRELIMINARY; PRT; 498 AA.
AC 09LZ54;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, last annotation update)
DE Hypothetical 55.5 kDa protein.
GN T2P11_70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxId-3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162971; CAB85985.1; -.
DR InterPro; IPR002068; HSP20.
DR InterPro; IPR000834; Zn_carboxypept.
DR PROSITE; PS00133; CARBOXYPEPT_2N_2; UNKNOWN_1.
DR PROSITE; PS01031; HSP20; 1.
KW Hypothetical protein.
SQ SEQUENCE 498 AA; 55530 MW; 0791437EB8817BE2 CRC64;

Query Match 41.9%; Score 44; DB 10; Length 498;
Best Local Similarity 70.0%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 5 GCFVGFDPDSRH 14
||| ||| |:
Db 132 GCFVGFDPDSRH 141

Search completed: March 27, 2003, 10:05:59
Job time : 13.6585 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:04:25 : Search time 4.40244 Seconds
(without alignments)
253.439 Million cell updates/sec

Title: US-09-675-208-2
Perfect score: 105
Sequence: 1 PVSVCFGVGFDPSEDSRH 19

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEM_PUB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEM_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEM_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEM_PUB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US09_NEM_PUB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US10_NEM_PUB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US60_NEM_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	45	42.9	296	10 US-09-873-880-34	Sequence 34, Appl
2	43	41.0	296	10 US-09-873-880-36	Sequence 36, Appl
3	42	40.0	382	9 US-09-738-626-4094	Sequence 4094, Ap
4	41	39.0	212	10 US-09-924-256A-22	Sequence 22, Appl
5	41	39.0	212	10 US-09-924-256A-34	Sequence 34, Appl
6	40	38.1	834	9 US-10-033-297-6	Sequence 6, Appl1
7	40	38.1	834	9 US-10-081-806-6	Sequence 6, Appl1
8	40	38.1	834	9 US-10-074-328-6	Sequence 6, Appl1
9	40	38.1	834	9 US-09-940-244-6	Sequence 6, Appl1
10	40	38.1	834	9 US-09-940-925A-6	Sequence 6, Appl1
11	40	38.1	836	9 US-09-940-244-259	Sequence 259, App
12	40	38.1	836	10 US-09-777-430A-8	Sequence 8, Appl1
13	40	38.1	836	10 US-09-777-430A-11	Sequence 11, Appl
14	40	38.1	842	10 US-09-777-430A-15	Sequence 15, Appl
15	40	38.1	842	10 US-09-777-430A-20	Sequence 20, Appl
16	40	38.1	842	10 US-09-777-430A-23	Sequence 23, Appl
17	40	38.1	842	10 US-09-777-430A-26	Sequence 26, Appl
18	39	37.1	228	9 US-09-813-153-110	Sequence 110, App
19	39	37.1	833	9 US-10-033-297-8	Sequence 8, Appl1

20	39	37.1	833	9	US-10-081-806-8	Sequence 8, Appl1
21	39	37.1	833	9	US-10-074-328-8	Sequence 8, Appl1
22	39	37.1	833	9	US-09-940-244-8	Sequence 8, Appl1
23	39	37.1	833	9	US-09-940-925A-8	Sequence 8, Appl1
24	38	36.2	143	10	US-09-894-998-11	Sequence 11, Appl
25	38	36.2	233	9	US-09-791-279-188	Sequence 188, App
26	38	36.2	284	9	US-09-853-450-20	Sequence 20, Appl1
27	38	36.2	369	10	US-09-838-955-3	Sequence 3, Appl1
28	38	36.2	440	9	US-09-738-626-6376	Sequence 6376, Ap
29	38	36.2	481	10	US-09-894-998-12	Sequence 12, Appl
30	38	36.2	831	9	US-10-033-297-5	Sequence 5, Appl1
31	38	36.2	831	9	US-10-081-806-5	Sequence 5, Appl1
32	38	36.2	831	9	US-10-074-328-5	Sequence 5, Appl1
33	38	36.2	831	9	US-09-940-244-5	Sequence 5, Appl1
34	38	36.2	831	9	US-09-940-925A-5	Sequence 5, Appl1
35	37.5	35.7	537	9	US-09-891-139A-12	Sequence 12, Appl
36	37.5	35.7	618	9	US-09-738-626-3975	Sequence 3975, Ap
37	37.5	35.7	722	9	US-10-128-870-23	Sequence 23, Appl
38	37.5	35.7	722	9	US-10-131-685-23	Sequence 23, Appl
39	37.5	35.7	844	10	US-09-813-148-4	Sequence 14, Appl
40	37.5	35.7	844	10	US-09-810-796-14	Sequence 10, Appl
41	37.5	35.7	871	9	US-10-128-870-20	Sequence 20, Appl
42	37.5	35.7	871	9	US-10-131-685-20	Sequence 20, Appl
43	37	35.2	132	9	US-09-975-036-6	Sequence 6, Appl1
44	37	35.2	358	10	US-09-864-761-48087	Sequence 48087, A
45	37	35.2	371	9	US-09-738-626-3981	Sequence 3981, Ap

ALIGNMENTS

RESULT 1

US-09-873-880-34
; Sequence 34, Application US/09873880

Patent No. US20020123118A1
; GENERAL INFORMATION:

APPLICANT: Sewalt, Vincent
APPLICANT: Falco, S. Carl

APPLICANT: Allen, Stephen M.
TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES

FILE REFERENCE: B81192 US CIP
CURRENT APPLICATION NUMBER: US/09/873.880

CURRENT FILING DATE: 2001-06-04
PRIOR APPLICATION NUMBER: 09/363.321

PRIOR FILING DATE: July 28, 1999
PRIOR APPLICATION NUMBER: 60/094.839

PRIOR FILING DATE: July 31, 1998
SOFTWARE: Microsoft Office 97

NUMBER OF SEQ ID NOS: 42
SEQ ID NO 34

LENGTH: 296
TYPE: PRT

ORGANISM: Zea mays
US-09-873-880-34

Query Match Score 45; DB 10; Length 296;
Best Local Similarity 50.0%; Pred. No. 12;

Matches 8; Conservative 4; Mismatches 4; Indels 0; Gaps 0;

QY 3 SVGCEVGFDPSEDSRH 18
:::1::111::111
Db 209 TLGEVAGFDPTEPTR 224

RESULT 2
US-09-873-880-36
; Sequence 36, Application US/09873880

Patent No. US20020123118A1
; GENERAL INFORMATION:

APPLICANT: Sewalt, Vincent
APPLICANT: Falco, S. Carl

APPLICANT: Allen, Stephen M.
TITLE OF INVENTION: GLYCINE METABOLISM ENZYMES

```
; FILE REFERENCE: Bb1192 US CIP
; CURRENT APPLICATION NUMBER: US/09/873,880
; CURRENT FILING DATE: 2001-06-04
; PRIOR APPLICATION NUMBER: 09/363,321
; PRIOR FILING DATE: July 28, 1999
; PRIOR APPLICATION NUMBER: 60/094,839
; PRIOR FILING DATE: July 31, 1998
; NUMBER OF SEQ ID NOS: 42
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 36
; LENGTH: 296
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-873-880-36

Query Match          41.0%; Score 43; DB 10; Length 296;
Best Local Similarity 57.1%; Pred. No. 24;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 5 GCFVGFDPDSR 18
1 : ||| : |||
Db 211 GEYAGFDPPTSR 224

RESULT 3
US-09-738-626-4094
; Sequence 4094, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKURO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentln ver. 3.0
; SEQ ID NO 4094
; LENGTH: 382
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-4094

Query Match          40.0%; Score 42; DB 9; Length 382;
Best Local Similarity 64.3%; Pred. No. 46;
Matches 9; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 2 VSVGCFVGFDPAS 15
1 : ||| || ||| ||
Db 249 VTVGTFVVDASSP 262

RESULT 4
US-09-924-256A-22
; Sequence 22, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
```

```
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 22
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Xanthoparmelia cumberlandia
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (136)
; OTHER INFORMATION: Xaa=unknown amino acid
US-09-924-256A-22

Query Match          39.0%; Score 41; DB 10; Length 212;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCVGFDPASPD 16
1 : ||| : ||| : |||
Db 22 PEDVGCYIGACATDYD 37

RESULT 5
US-09-924-256A-34
; Sequence 34, Application US/09924256A
; Patent No. US20020127659A1
; GENERAL INFORMATION:
; APPLICANT: Waters, Barbara
; APPLICANT: Miao, Vivian
; APPLICANT: Ho, Yap
; APPLICANT: Tong, Seow
; TITLE OF INVENTION: METHOD FOR ISOLATION OF BIOSYNTHESIS GENES FOR
; TITLE OF INVENTION: BIOACTIVE MOLECULES
; FILE REFERENCE: 9993-006
; CURRENT APPLICATION NUMBER: US/09/924,256A
; CURRENT FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 08/861,774
; PRIOR FILING DATE: 2001-04-13
; NUMBER OF SEQ ID NOS: 94
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 34
; LENGTH: 212
; TYPE: PRT
; ORGANISM: Leptogium corniculatum
US-09-924-256A-34

Query Match          39.0%; Score 41; DB 10; Length 212;
Best Local Similarity 43.8%; Pred. No. 36;
Matches 7; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 1 PVSVCVGFDPASPD 16
1 : ||| : ||| : |||
Db 22 PEDVGCYIGACATDYD 37

RESULT 6
US-10-033-297-6
; Sequence 6, Application US/10033297
; Publication No. US20020187486A1
; GENERAL INFORMATION:
; APPLICANT: Hall, Jeff G.
; APPLICANT: Lyamichev, Victor I.
; APPLICANT: Mast, Andrea L.
; APPLICANT: Brow, Mary Ann D.
; TITLE OF INVENTION: Detection Of Nucleic Acids By Multiple
```

Sequential Invasive Cleavages
NUMBER OF SEQUENCES: 163
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medien & Carroll, LLP
STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/033,297
FILING DATE: 12-NOV-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/350,597
FILING DATE: 09-JUL-1999
APPLICATION NUMBER: US/08/823,516
FILING DATE: 24-MAR-1997
APPLICATION NUMBER: PCT/US97/01072
FILING DATE: 21-JAN-1997
APPLICATION NUMBER: US 08/759,038
FILING DATE: 02-DEC-1996
APPLICATION NUMBER: US 08/758,314
FILING DATE: 02-DEC-1996
APPLICATION NUMBER: US 08/756,386
FILING DATE: 29-NOV-1996
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02736
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-033-297-6
Query Match 38.1%; Score 40; DB 9; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 FVGFDAPEPDSRH 19
DB 64 FVFDKAKAPSRH 76

RESULT 7
US-10-081-806-6
Sequence 6, Application US/10081806
Publication No. US20020197623A1
GENERAL INFORMATION:
APPLICANT: Prudent, James R.
Hail, Jeff G.
Lyamichev, Victor I.
TITLE OF INVENTION: Invasive Cleavage of Nucleic Acids
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Medien & Carroll, LLP

STREET: 220 Montgomery Street, Suite 2200
CITY: San Francisco
STATE: California
COUNTRY: United States of America
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/081,806
FILING DATE: 22-Feb-2002
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/756,386
FILING DATE: <unknown>
APPLICATION NUMBER: US 08/682,853
FILING DATE: 12-JUL-1996
APPLICATION NUMBER: US 08/599,491
FILING DATE: 24-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ingolia, Diane E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: FORS-02564
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-081-806-6
Query Match 38.1%; Score 40; DB 9; Length 834;
Best Local Similarity 61.5%; Pred. No. 2.2e+02;
Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

OY 7 FVGFDAPEPDSRH 19
DB 64 FVFDKAKAPSRH 76

RESULT 8
US-10-074-328-6
Sequence 6, Application US/10074328
Publication No. US20030013098A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
GROTELUESCHEN HALL, JEFF S.
OLIVE, DAVID M.
PRUDENT, JAMES R.
TITLE OF INVENTION: INVASION OF NUCLEIC ACID SEQUENCES BY
INVADER-DIRECTED CLEAVAGE
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESSES:
ADDRESSEE: MEDIEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/074,328
FILING DATE: 12-Feb-2002
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/599,491
FILING DATE: 23-Jan-1996
ATTORNEY/AGENT INFORMATION:
NAME: INGOLTA, DIANE E.
REGISTRATION NUMBER: P-40,027
REFERENCE/DOCKET NUMBER: FORS-01802
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-074-328-6

Query Match 38.1% Score 40: DB 9: Length 834;
Best Local Similarity 61.5% Pred. No. 2.2e+02;
Matches 8: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
11111111
DB 64 FVFEADAKAPSRH 76

RESULT 9
US-09-940-244-6
Sequence 6, Application US/09940244
Publication No. US20030044796A1
GENERAL INFORMATION:
APPLICANT: Neri, Bruce P.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamichev, Victor
APPLICANT: Smith, Lloyd M.
TITLE OF INVENTION: Reactions on Dendrimers
FILE REFERENCE: FORS-06478
CURRENT APPLICATION NUMBER: US/09/940,244
CURRENT FILING DATE: 2002-05-06
NUMBER OF SEQ ID NOS: 422
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 834
TYPE: PRT
ORGANISM: Thermus thermophilus
US-09-940-244-6

Query Match 38.1% Score 40: DB 9: Length 834;
Best Local Similarity 61.5% Pred. No. 2.2e+02;
Matches 8: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
11111111
DB 64 FVFEADAKAPSRH 76

RESULT 10
US-09-940-925A-6
Sequence 6, Application US/09940925A
Publication No. US20030054338A1
GENERAL INFORMATION:
APPLICANT: BROW, MARY ANN D.
APPLICANT: LYAMICHEV, VICTOR I.
APPLICANT: OLIVE, DAVID M.
TITLE OF INVENTION: RAPID DETECTION AND IDENTIFICATION OF
PATHOGENS
NUMBER OF SEQUENCES: 165

CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: UNITED STATES OF AMERICA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/940,925A
FILING DATE: 10-Jun-2002
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: CARROLL, PETER G.
REGISTRATION NUMBER: 32,837
REFERENCE/DOCKET NUMBER: FORS-01756
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 834 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-940-925A-6

Query Match 38.1% Score 40: DB 9: Length 834;
Best Local Similarity 61.5% Pred. No. 2.2e+02;
Matches 8: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
11111111
DB 64 FVFEADAKAPSRH 76

RESULT 11
US-09-940-244-259
Sequence 259, Application US/09940244
Publication No. US20030044796A1
GENERAL INFORMATION:
APPLICANT: Neri, Bruce P.
APPLICANT: Hall, Jeff G.
APPLICANT: Lyamichev, Victor
APPLICANT: Smith, Lloyd M.
TITLE OF INVENTION: Reactions on Dendrimers
FILE REFERENCE: FORS-06478
CURRENT APPLICATION NUMBER: US/09/940,244
CURRENT FILING DATE: 2002-05-06
NUMBER OF SEQ ID NOS: 422
SOFTWARE: PatentIn version 3.1
SEQ ID NO 259
LENGTH: 836
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic
US-09-940-244-259

Query Match 38.1% Score 40: DB 9: Length 836;
Best Local Similarity 61.5% Pred. No. 2.2e+02;
Matches 8: Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 7 FVGFDAEPDSRH 19
11111111
DB 66 FVFEADAKAPSRH 78

```
RESULT 12
US-09-777-430A-8
; Sequence 8, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; TITLE OF INVENTION: Charge Tags and the Separation of Nucleic Acid Molecules
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-8
```

```
Query Match          38.1%  Score 40;  DB 10;  Length 836;
Best Local Similarity 61.5%  Pred. No. 2.2e+02;
Matches 8;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
DB 66 FVFDKAPSPFRH 78
```

```
RESULT 13
US-09-777-430A-11
; Sequence 11, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 836
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-11
```

```
Query Match          38.1%  Score 40;  DB 10;  Length 836;
Best Local Similarity 61.5%  Pred. No. 2.2e+02;
Matches 8;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
DB 66 FVFDKAPSPFRH 78
```

```
RESULT 14
US-09-777-430A-15
; Sequence 15, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
```

```
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; TITLE OF INVENTION: Charge Tags and the Separation of Nucleic Acid Molecules
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 15
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-15
```

```
Query Match          38.1%  Score 40;  DB 10;  Length 842;
Best Local Similarity 61.5%  Pred. No. 2.2e+02;
Matches 8;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
DB 66 FVFDKAPSPFRH 78
```

```
RESULT 15
US-09-777-430A-20
; Sequence 20, Application US/09777430A
; Patent No. US20020128465A1
; GENERAL INFORMATION:
; APPLICANT: Lyamichev, Victor
; APPLICANT: Skrzypczynski, Zbigniew
; APPLICANT: Allawi, Halim T.
; APPLICANT: Wayland, Sarah R.
; APPLICANT: Takova, Tsetska
; APPLICANT: Neir, Bruce P.
; TITLE OF INVENTION: Charge Tags and the Separation of Nucleic Acid Molecules
; FILE REFERENCE: FORS-04912
; CURRENT APPLICATION NUMBER: US/09/777,430A
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 85
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 20
; LENGTH: 842
; TYPE: PRT
; ORGANISM: Thermus thermophilus
US-09-777-430A-20
```

```
Query Match          38.1%  Score 40;  DB 10;  Length 842;
Best Local Similarity 61.5%  Pred. No. 2.2e+02;
Matches 8;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
QY 7 FVGFDASEPPDSRH 19
DB 66 FVFDKAPSPFRH 78
```

```
Search completed: March 27, 2003, 10:20:36
Job time : 5.40244 secs
```


GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:40 ; Search time 7.31707 Seconds
(without alignments)
120.634 Million cell updates/sec

Title: US-09-675-208-1
Perfect score: 181
Sequence: 1 FGMCTWDAFYLTVPQGVIEGVRLVDDGC 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued_Patents_AA.*
- 2: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
- 3: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
- 4: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
- 5: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
- 6: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
- 7: /cgn2_6/ptodata/1/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	30	4	US-08-846-234-1
2	181	100.0	784	4	US-08-846-234-5
3	60	33.1	132	1	US-08-100-744-4
4	60	33.1	132	1	US-08-284-784-4
5	60	33.1	132	2	US-08-854-811-4
6	60	33.1	140	5	PCT-US90-06962-1
7	60	33.1	145	1	US-07-830-330-1
8	60	33.1	145	1	US-08-187-780-3
9	60	33.1	145	2	US-08-478-485-3
10	60	33.1	145	2	US-08-478-486F-3
11	60	33.1	146	2	US-08-231-894A-11
12	60	33.1	146	2	US-08-231-894A-12
13	60	33.1	146	2	US-08-231-894A-13
14	60	33.1	146	2	US-08-231-894A-14
15	60	33.1	146	2	US-08-231-894A-15
16	60	33.1	146	3	US-09-105-678A-15
17	60	33.1	146	3	US-09-421-208-49
18	60	33.1	146	4	US-09-385-114-2
19	60	33.1	146	4	US-09-417-721-3
20	60	33.1	146	4	US-09-417-721-5
21	60	33.1	146	6	5464943-6
22	60	33.1	146	6	5464943-8
23	60	33.1	146	6	5464943-10
24	60	33.1	146	6	5464943-12
25	60	33.1	146	6	5464943-14
26	60	33.1	146	6	5464943-25
27	60	33.1	146	6	5464943-26

28	60	33.1	147	6	5175147-8	Patent No. 5175147
29	60	33.1	147	6	5314872-1	Patent No. 5314872
30	60	33.1	150	1	US-08-441-629-8	Sequence 8, Appl1
31	60	33.1	150	3	US-08-776-207-8	Sequence 8, Appl1
32	60	33.1	150	3	US-09-507-773-8	Sequence 8, Appl1
33	60	33.1	150	5	PCT-US95-09172-8	Sequence 8, Appl1
34	60	33.1	153	3	US-08-325-186-2	Sequence 2, Appl1
35	60	33.1	154	2	US-08-438-439C-24	Sequence 24, Appl1
36	60	33.1	154	3	US-08-325-186-1	Sequence 1, Appl1
37	60	33.1	154	5	PCT-US91-02186-6	Sequence 6, Appl1
38	60	33.1	155	1	US-07-959-369-6	Sequence 6, Appl1
39	60	33.1	155	1	US-07-959-369-7	Sequence 7, Appl1
40	60	33.1	155	1	US-08-023-757-2	Sequence 2, Appl1
41	60	33.1	155	1	US-08-023-757-4	Sequence 4, Appl1
42	60	33.1	155	1	US-07-842-177A-1	Sequence 1, Appl1
43	60	33.1	155	1	US-08-177-502-2	Sequence 2, Appl1
44	60	33.1	155	1	US-08-177-502-4	Sequence 4, Appl1
45	60	33.1	155	1	US-08-439-725A-10	Sequence 10, Appl1

ALIGNMENTS

RESULT 1
US-08-846-234-1
Sequence 1, Application US/08846234
Patent No. 6166292

GENERAL INFORMATION:
APPLICANT: OSUMI Chleko
APPLICANT: NOZAKI Jinshi

APPLICANT: KIDA Takao
TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT

NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESSES:

ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR

CITY: ARLINGTON
STATE: VIRGINIA

COUNTRY: USA
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/846,234
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618
TELEPHONE: (703)-413-3000

TELEFAX: (703)-413-2220
INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 30 amino acids

TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: peptide
FRAGMENT TYPE: internal

US-08-846-234-1

Query Match

Best Local Similarity 100.0%; Score 181, DB 4;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 FGMCTWDAFYLTVPQGVIEGVRLVDDGC 30

1 FGMCTWDAFYLTVPQGVIEGVRLVDDGC 30

DB

RESULT 2
US-08-846-234-5
; Sequence 5, Application US/08846234
; Patent No. 6166292
; GENERAL INFORMATION:
; APPLICANT: OSUMI Chieko
; APPLICANT: NOZAKI Jinshi
; APPLICANT: KIDA Takao
; TITLE OF INVENTION: RAFFINOSE SYNTHASE GENE, METHOD FOR
; TITLE OF INVENTION: PRODUCING RAFFINOSE, AND TRANSGENIC PLANT
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR
; CITY: ARLINGTON
; STATE: VIRGINIA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/846,234
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: NORMAN F. OBLON
; REGISTRATION NUMBER: 24,618
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)-413-3000
; TELEFAX: (703)-413-2220
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 784 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-846-234-5
Query Match 100.0%; Score 181; DB 4; Length 784;
Best Local Similarity 100.0%; Pred. No. 2,7e-18;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 FGCWTDAFYLVHPGVIEGVRLVDDGC 30
Db 215 FGCWTDAFYLVHPGVIEGVRLVDDGC 244
RESULT 3
US-08-100-744-4
; Sequence 4, Application US/08100744
; Patent No. 5563046
; GENERAL INFORMATION:
; APPLICANT: MASCARENHAS, DESMOND
; APPLICANT: ZHANG, SUNNY
; APPLICANT: OLSEN, PAMELA
; APPLICANT: OLSEN, PAMELA
; APPLICANT: CARRILLO, PEDRO A.
; TITLE OF INVENTION: POLYPEPTIDE FUSIONS TO
; TITLE OF INVENTION: INTERLEUKIN-1-LIKE POLYPEPTIDES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,784
; FILING DATE: 02-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: PARK, FREDIE K.
; REGISTRATION NUMBER: 35,636
; REFERENCE/DOCKET NUMBER: 22095-20275.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-100-744-4

SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/100,744
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LUTHER, BARBARA J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: 22095-20275.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-100-744-4
Query Match 33.1%; Score 60; DB 1; Length 132;
Best Local Similarity 40.0%; Pred. No. 0.17;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
Qy 3 WCTWDAFYLVHPGVIEGVRLVDD 27
Db 10 YCKNGGFRLRHDPGRVDCGRKSD 34
RESULT 4
US-08-284-784-4
; Sequence 4, Application US/08284784
; Patent No. 5629172
; GENERAL INFORMATION:
; APPLICANT: MASCARENHAS, DESMOND
; APPLICANT: ZHANG, YANG
; APPLICANT: OLSEN, PAMELA S.
; APPLICANT: OLSEN, PAMELA R.
; APPLICANT: CARRILLO, PEDRO A.
; TITLE OF INVENTION: EXPRESSION OF FUSION POLYPEPTIDES
; TITLE OF INVENTION: TRANSPORTED OUT OF THE CYTOPLASM WITHOUT LEADER SEQUENCES
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 Page Mill Road
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/284,784
; FILING DATE: 02-AUG-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: PARK, FREDIE K.
; REGISTRATION NUMBER: 35,636
; REFERENCE/DOCKET NUMBER: 22095-20275.20
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 813-5600
; TELEFAX: (415) 494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 132 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

APPLICATION NUMBER: 07/806,771
FILING DATE: December 6, 1991
APPLICATION NUMBER: 07/177,506
FILING DATE: April 4, 1988
APPLICATION NUMBER: 07/062,925
FILING DATE: June 16, 1987
ATTORNEY/AGENT INFORMATION:
NAME: Joseph R. Robinson
REGISTRATION NUMBER: 33,448
REFERENCE/DOCKET NUMBER: 5986/13586-US6
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 145
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
DESCRIPTION: Protein
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: This sequence,
OTHER INFORMATION: corresponding to bovine basic fibroblast
OTHER INFORMATION: growth factor, can be found in Table 1,
OTHER INFORMATION: page 9, lines 9, 14, and 19, in the
OTHER INFORMATION: application, as filed.
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO: 1-145
US-08-478-485-3
Query Match 33.1%; Score 60; DB 2; Length 145;
Best Local Similarity 40.0%; Pred. No. 0.19;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
OY 3 WCTMDAFYLVHPQGVIEGVRLVD 27
Db 24 YCKNGFRLRHDPGRVDGVREKSD 48
RESULT 10
US-08-478-486F-3
Sequence 3, Application US/08478486F
Patent No. 6432702
GENERAL INFORMATION:
APPLICANT: CLAUDIO BASILICO
APPLICANT: DANIELA TALARICO
TITLE OF INVENTION: MAMMALIAN GROWTH FACTOR
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Darby & Darby P.C.
STREET: 805 Third Avenue
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10022
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy Diskette, 3+ inch,
MEDIUM TYPE: 1.44 MB storage
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC/MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,486F
FILING DATE: June 7, 1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/187,780
FILING DATE: January 25, 1994
APPLICATION NUMBER: 07/901,705
FILING DATE: June 22, 1992
APPLICATION NUMBER: 07/806,771
FILING DATE: December 6, 1991
APPLICATION NUMBER: 07/177,506
FILING DATE: April 4, 1988
APPLICATION NUMBER: 07/062,925
FILING DATE: June 16, 1987
ATTORNEY/AGENT INFORMATION:
NAME: Howard M. Frankfort
REGISTRATION NUMBER: 32,613
REFERENCE/DOCKET NUMBER: 5986/13586-US7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 527-7700
TELEFAX: (212) 753-6237
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 145
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Protein
DESCRIPTION: Protein
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION: This sequence,
OTHER INFORMATION: corresponding to bovine basic fibroblast
OTHER INFORMATION: growth factor, can be found in Table 1,
OTHER INFORMATION: page 9, lines 9, 14, and 19, in the
OTHER INFORMATION: application, as filed.
US-08-478-486F-3
Query Match 33.1%; Score 60; DB 4; Length 145;
Best Local Similarity 40.0%; Pred. No. 0.19;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
OY 3 WCTMDAFYLVHPQGVIEGVRLVD 27
Db 24 YCKNGFRLRHDPGRVDGVREKSD 48
RESULT 11
US-08-231-894A-11
Sequence 11, Application US/08231894A
Patent No. 5851990
GENERAL INFORMATION:
APPLICANT: FUJISHIMA, AKIRA
APPLICANT: FUKUDA, TSUNEHIKO
TITLE OF INVENTION: BRGF PROTEIN AND ITS PRODUCTION
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: DAVID G. COMLIN; DIKE, BRONSTEIN, ROBERTS
STREET: 130 WATER STREET
CITY: BOSTON
STATE: MASSACHUSETTS
COUNTRY: US
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS

Fri Apr 4 12:31:19 2003

us-09-675-208-1.rai

Page 8

ORIGINAL SOURCE:
US-08-231-894A-15

Query Match	33.1%	Score 60;	DB 2;	Length 146;
Best Local Similarity	40.0%;	Pred. No. 0.19;		
Matches 10;	Conservative 5;	Mismatches 10;	Indels 0;	Gaps 0;

```
Oy      3 WCITDAFYLTVHPQGVIEGVRHLVD 27
          :| |::||| |
Db     24 YCKNGGFELRIHPPDGRVDGVRKESD 48
```

Search completed: March 27, 2003, 10:07:33
Job time : 8.31707 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 21.2195 Seconds

(without alignments)
188.389 Million cell updates/sec

Title: US-09-675-208-1

Perfect score: 181

Sequence: 1 FGWCTWDAPFLYLVHPOGVIEGVRLVDGCG 30

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_101002.*

1: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	30	19	AAW53567
2	181	100.0	30	20	AAV17422
3	181	100.0	784	19	AAW53570
4	181	100.0	784	20	AAV17417
5	158	87.3	758	21	AAV70978
6	148	81.8	799	19	AAW57886
7	148	81.8	799	20	AAV30142
8	146	80.7	587	19	AAW57888
9	146	80.7	587	20	AAV30144
10	146	80.7	780	22	AAW57859

11	146	80.7	781	19	AAW57887
12	146	80.7	781	20	AAV30143
13	146	80.7	781	22	AAW57888
14	144	79.6	572	20	AAV32075
15	144	79.6	572	20	AAV32075
16	144	79.6	777	20	AAV32074
17	144	79.6	783	23	AAW57886
18	122	67.4	783	20	AAV32073
19	107	59.7	751	21	AAV70980
20	106	58.6	750	20	AAV70977
21	106	58.6	750	20	AAV70977
22	102	56.4	807	23	AAW57886
23	98	54.1	770	21	AAV70976
24	97	53.6	841	21	AAV70981
25	71	39.2	756	21	AAV70975
26	68	37.6	150	11	AAW03961
27	62	34.3	40	14	AAW43287
28	60	33.1	40	14	AAW43286
29	60	33.1	43	13	AAW4342
30	60	33.1	44	21	AAW18542
31	60	33.1	45	8	AAW71559
32	60	33.1	45	14	AAW43278
33	60	33.1	45	21	AAW18551
34	60	33.1	86	9	AAW81933
35	60	33.1	101	10	AAW90557
36	60	33.1	105	10	AAW90558
37	60	33.1	114	10	AAW90559
38	60	33.1	114	22	AAW60696
39	60	33.1	118	10	AAW90560
40	60	33.1	121	23	AAW81292
41	60	33.1	123	9	AAW90561
42	60	33.1	129	10	AAW81940
43	60	33.1	129	10	AAW90562
44	60	33.1	129	10	AAW90564
45	60	33.1	132	20	AAV17995

ALIGNMENTS

RESULT 1	AAW53567	standard; peptide: 30 AA.
ID	AAW53567	
AC	AAW53567	
DF	06-JUL-1998	(first entry)
DE	Cucurbituraculose synthase residues 215 to 244.	
KW	Cucurbituraculose synthase; sucrose; galactinol.	
OS	Cucurbituraculose synthase.	
PN	JPI0084973-A.	
PD	07-APR-1998.	
PF	28-APR-1997;	97JP-011124.
PR	26-JUL-1996;	96JP-0198079.
PR	26-APR-1996;	96JP-0107682.
PA	(AJIN) AJINOMOTO KK.	
DR	WPI: 1998-264858/24.	
PT	Raffinose synthase gene - useful for preparation of raffinose in	
PT	transformed plant	
PS	Claim 2: Page 16: 26pp; Japanese.	
CC	The present sequence is a cucurbituraculose synthase fragment.	

Soybean raffinose
Amino acid sequenc
Soybean raffinose
Rapeseed raffinose
Mustard raffinose
Herbicideally activ
Sugarbeet raffinose
Wheat raffinose sy
Rice raffinose syn
Soybean raffinose
Herbicideally activ
Rice raffinose syn
Wheat raffinose sy
Corn raffinose syn
Soybean raffinose
Basic fibroblast g
ECF antagonist [AI
ECF antagonist bfg
ECF residues 27-69
Immunogenic peptid
Fibroblast Growth
ECF antagonist bfg
Immunogenic peptid
Human basic fibrob
rhbFGF mutlein CS10
rhbFGF mutlein CS10
rhbFGF mutlein CS14
Human basic fibrob
rhbFGF mutlein C118
Human FGF2 core st
rhbFGF mutlein C123
Human basic fibrob
rhbFGF mutlein C129
rhbFGF mutlein CS23
Human basic fibrob

CC Raffinose synthase forms raffinose from sucrose and galactinol, has an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees C, has a molecular weight of 75 to 95 kDa by gel filtration or 90 kDa by PAGE and SDS-PAGE under reductive conditions and is inhibited by iodacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 30 AA;

Query Match 100.0%; Score 181; DB 19; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30
|||||
DB 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30

RESULT 2
AAV17422
ID AAV17422 standard; peptide; 30 AA.

XX AAV17422;

XX 29-JUL-1999 (first entry)

DE Cucurbit raffinose synthase peptide SEQ ID NO:1.

KW Raffinose synthase; sucrose; galactinol.

OS Cucumis sativus.

PN JP1123080-A.

PD 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

PA (AJIN) AJINOMOTO KK.

DR WPI; 1999-340516/29.

PT New raffinose synthase gene - for production of raffinose from sucrose and galactinol

PS Example 2; Page 21; 37pp; Japanese.

CC The present invention describes a raffinose synthase, having an activity of forming raffinose from sucrose and galactinol. The raffinose synthase gene can be used for expression in a plant for the production of raffinose. The raffinose synthase can give raffinose from sucrose and galactinol efficiently. The present sequence represents a raffinose synthase peptide from cucumber.

XX Sequence 30 AA;

Query Match 100.0%; Score 181; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 2.5e-19;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30
|||||
DB 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30

RESULT 3
AAW53570

ID AAW53570 standard; Protein; 784 AA.

XX AAW53570;

DT 06-JUL-1998 (first entry)

XX Cucurbit raffinose synthase.

KW Cucurbit; raffinose synthase; sucrose; galactinol.

OS Cucumis sativus.

PN JP10084973-A.

PD 07-APR-1998.

PF 28-APR-1997; 97JP-0111124.

PR 26-JUL-1996; 96JP-0198079.

PR 26-APR-1996; 96JP-0107682.

PA (AJIN) AJINOMOTO KK.

DR WPI; 1998-264858/24.

DR N-PSDB; AAV22250.

PT Raffinose synthase gene - useful for preparation of raffinose in transformed plant

PS Claim 3; Pages 17-20; 26pp; Japanese.

CC The present sequence is cucumber raffinose synthase, which forms raffinose from sucrose and galactinol, has an optimum pH of 6 to 8 and working temperature of 35 to 40 degrees C, has a molecular weight of 75 to 95 kDa by gel filtration or 90 to 100 kDa by PAGE and SDS-PAGE under reductive conditions and is inhibited by iodacetamide, N-ethylmaleimide and myoinositol.

XX Sequence 784 AA;

Query Match 100.0%; Score 181; DB 19; Length 784;
Best Local Similarity 100.0%; Pred. No. 1.2e-17;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 30
|||||
DB 215 FGMCTWDAFYLTVHPGVIEGVRLHVDGCG 244

RESULT 4
AAV17417

ID AAV17417 standard; Protein; 784 AA.

XX AAV17417;

DT 29-JUL-1999 (first entry)

DE Cucurbit raffinose synthase.

KW Raffinose synthase; sucrose; galactinol.

OS Cucumis sativus.

PN JP1123080-A.

PD 11-MAY-1999.

PF 24-OCT-1997; 97JP-0292969.

PR 24-OCT-1997; 97JP-0292969.

PA (AJIN) AJINOMOTO KK.

DR WPI; 1999-340516/29.

DR N-PSDB; AAX61238.

PT New raffinose synthase gene - for production of raffinose from sucrose and galactinol

XX Claim 2; Page 25-27; 37pp; Japanese.
PS
XX
CC The present invention describes a raffinose synthase, having an activity
CC of forming raffinose from sucrose and galactinol. The raffinose synthase
CC gene can be used for expression in a plant for the production of
CC raffinose. The raffinose synthase can give raffinose from sucrose
CC and galactinol efficiently. The present sequence represents raffinose
CC synthase from cucumber.
XX
SQ Sequence 784 AA:

Query Match 100.0%; Score 181; DB 20; Length 784;
Best Local Similarity 100.0%; Pred. No. 1,2e-17;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLVHPGVIEGVRHLVDGCG 30
DB 215 FGMCTWDAFYLVHPGVIEGVRHLVDGCG 244
|||||

RESULT 5
AAV70978
ID AAV70978 standard; Protein; 758 AA.
XX
AC AAV70978;
XX
DT 09-AUG-2000 (first entry)
XX
DE Soybean raffinose synthase from clone sfl1.pk125.d4.
XX
KW Soybean; raffinose synthase; raffinose saccharide;
KW clone sfl1.pk125.d4; nutritional; soy protein.
XX
OS Glycine max.
XX
PM WO200024915-A2.
PN
XX
PD 04-MAY-2000.
XX
PF 22-OCT-1999; 99MO-US24923.
XX
PR 23-OCT-1998; 98US-0105451.
XX
PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Allen SM, Hiltz WD;
XX
DR WPI: 2000-350754/30.
DR N-PSDB: AAD00335.
XX
XX
PT Nucleic acids and encoded proteins involved in the biosynthesis of
PT raffinose, useful for producing soybean seeds with a reduced raffinose
PT content and therefore improved nutritional quality -
XX
PS Claim 2; Page 47-49; 58pp; English.
XX
XX The present sequence is a raffinose synthase from
CC clone sfl1.pk125.d4 isolated from a soybean immature flower cDNA
CC library sfl1. Raffinose synthase is involved in the biosynthesis
CC of raffinose and higher homologues in the raffinose saccharide family
CC from sucrose. The present sequence is useful for reducing the raffinose
CC saccharide content of soybean seeds which improves the nutritional
CC quality of the soy protein products derived from them.
XX
SQ Sequence 758 AA:

Query Match 87.3%; Score 158; DB 21; Length 758;
Best Local Similarity 86.7%; Pred. No. 2,6e-14;
Matches 26; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLVHPGVIEGVRHLVDGCG 30
|||||

DB 202 FGMCTWDAFYLVHPGVIEGVRGKGLVDGCG 231

RESULT 6
AAW57886
ID AAW57886 standard; Protein; 799 AA.
XX
AC AAW57886;
XX
DT 23-SEP-1998 (first entry)
XX
DE Broad bean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; broad bean.
XX
OS Vicia faba.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Wantanabe E;
XX
DR WPI: 1998-324670/29.
DR N-PSDB: AAV40800.
XX
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
PS Claim 1; Page 26-29; 44pp; English.
XX
XX
CC This sequence represents the broad bean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 799 AA:

Query Match 81.8%; Score 148; DB 19; Length 799;
Best Local Similarity 80.0%; Pred. No. 8,1e-13;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLVHPGVIEGVRHLVDGCG 30
DB 238 FGMCTWDAFYLVHPGVIEGVRKSLVDGCG 267
|||||

RESULT 7
AAV30142
ID AAV30142 standard; Protein; 799 AA.
XX
AC AAV30142;
XX
DT 26-OCT-1999 (first entry)
XX
DE Amino acid sequence of a raffinose synthase protein.
XX
KW Raffinose synthase; plant; broad bean; sucrose; raffinose.
XX
OS Vicia faba.

```

XX  JP11215984-A.
XX
XX  10-AUG-1999.
XX
XX  12-DEC-1997; 97JP-0342899.
XX
XX  28-NOV-1997; 97JP-0329006.
XX
XX  18-DEC-1996; 96JP-0338673.
XX
XX  (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX  WPI; 1999-51112/43.
XX
XX  N-PSDB; AA210001.
XX
XX  New raffinose synthase gene - is prepared from a plant material
XX
XX  Claim 5; Page 19-21; 40pp; Japanese.
XX
XX  The present sequence represents a raffinose synthase protein. The
XX  sequence is isolated from plant material of broad beans. The
XX  protein forms raffinose by complexing alpha(1 to 6) - D-galactosyl
XX  hydroxyl group of the 6C of D-glucose residue in sucrose molecules.
XX
XX  Sequence 799 AA;

Query Match      81.8%; Score 148; DB 20; Length 799;
Best Local Similarity 80.0%; Pred. No. 8.1e-13;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  1 FGCWTDAPFLYVHPQVIEGVRHLYDGGC 30
    |||||
DB  238 FGCWTDAPFLYVHPQVIEGVRHLYDGGC 267

RESULT 8
AAW57888
ID  AAW57888 standard; Protein; 587 AA.
XX
XX  AAW57888;
XX
XX  23-SEP-1998 (first entry)
XX
XX  Japanese artichoke raffinose synthetase.
XX
XX  Raffinose synthetase; metabolism modification; food additive;
XX  gastrointestinal flora; Japanese artichoke.
XX
XX  Stachys sieboldii.
XX
XX  EP849359-A2.
XX
XX  24-JUN-1998.
XX
XX  18-DEC-1997; 97EP-0122417.
XX
XX  18-DEC-1996; 96JP-0338673.
XX
XX  (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX  Oeda K, Wantanabe E;
XX
XX  WPI; 1998-324670/29.
XX
XX  N-PSDB; AAV40802.
XX
XX  New nucleic acid molecule encoding plant raffinose synthetase -
XX  capable of producing raffinose, used as food additives with
XX  beneficial effects on gastrointestinal flora
XX
XX  Claim 1; Page 36-38; 44pp; English.
XX
XX  This sequence is the Japanese artichoke raffinose synthetase of the
XX  invention. The raffinose synthetase is capable of producing raffinose by

```

```

CC  combining a D-galactosyl group through an alpha (1-6) bond with a
CC  hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC  residue in a sucrose molecule. The DNA can be used to modify metabolism
CC  of a host organism by introducing into the host organism or its cell so
CC  that the content of the raffinose family oligosaccharides in the host
CC  organism or cell is changed. Raffinose oligosaccharides are useful as
CC  food additives with beneficial effects on the gastrointestinal flora.
XX
XX  Sequence 587 AA;

Query Match      80.7%; Score 146; DB 19; Length 587;
Best Local Similarity 80.0%; Pred. No. 1.1e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  1 FGCWTDAPFLYVHPQVIEGVRHLYDGGC 30
    |||||
DB  107 FGCWTDAPFLYVHPQVIEGVRHLYDGGC 136

RESULT 9
AA30144
ID  AA30144 standard; Protein; 587 AA.
XX
XX  AA30144;
XX
XX  26-OCT-1999 (first entry)
XX
XX  Amino acid sequence of a raffinose synthase protein.
XX
XX  Raffinose synthase; plant; sucrose; raffinose.
XX
XX  Stachys sieboldii.
XX
XX  JP11215984-A.
XX
XX  10-AUG-1999.
XX
XX  12-DEC-1997; 97JP-0342899.
XX
XX  28-NOV-1997; 97JP-0329006.
XX
XX  18-DEC-1996; 96JP-0338673.
XX
XX  (SUMO ) SUMITOMO CHEM CO LTD.
XX
XX  WPI; 1999-51112/43.
XX
XX  N-PSDB; AA210003.
XX
XX  New raffinose synthase gene - is prepared from a plant material
XX
XX  Claim 12; Page 30-31; 40pp; Japanese.
XX
XX  The present sequence represents a raffinose synthase protein. The
XX  sequence is isolated from plant material. The protein forms raffinose
XX  by complexing alpha(1 to 6) - D-galactosyl hydroxyl group of the 6C
XX  of D-glucose residue in sucrose molecules.
XX
XX  Sequence 587 AA;

Query Match      80.7%; Score 146; DB 20; Length 587;
Best Local Similarity 80.0%; Pred. No. 1.1e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY  1 FGCWTDAPFLYVHPQVIEGVRHLYDGGC 30
    |||||
DB  107 FGCWTDAPFLYVHPQVIEGVRHLYDGGC 136

RESULT 10
AAB98659
ID  AAB98659 standard; Protein; 780 AA.
XX
XX  AAB98659;
XX

```

DT 17-AUG-2001 (first entry)
XX
DE Soybean protein: SEQ ID 1.
XX
KW Mutant; mutain; raffinose synthase; raffinose oligosaccharide reduction;
KW plant; soybean.
XX
OS Glycine max.
XX
PN JP2001078783-A.
XX
PD 27-MAR-2001.
XX
PF 03-JUL-2000; 2000JP-0200571.
XX
PR 09-JUL-1999; 99JP-0196036.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
XX WPI; 2001-313373/33.
DR N-PSDB; AAH27438.
XX
PT Novel mutant protein of raffinose synthase is useful for reducing the
PT raffinose oligosaccharide content in a plant body -
PS Disclosure; Page 18-20; 30pp; Japanese.
XX
CC The present invention relates to a mutant protein of raffinose synthase
CC in which at least one aromatic amino acid present at the position of
CC about 1-7 amino acids from the N-terminus is deleted or replaced. The
CC mutant protein can be used for reducing the raffinose oligosaccharide
CC content in a plant body. The present protein from soybean, was used in
CC the present invention.
XX
SQ Sequence 780 AA:

Query Match 80.7%; Score 146; DB 22; Length 780;
Best Local Similarity 80.0%; Pred. No. 1.5e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 30
DB 220 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 249

RESULT 11
ID AAW57887 standard; Protein: 781 AA.
XX
AC AAW57887;
XX
DT 23-SEP-1998 (first entry)
XX
DE Soybean raffinose synthetase.
XX
KW Raffinose synthetase; metabolism modification; food additive;
KW gastrointestinal flora; soybean.
XX
OS Glycine max.
XX
PN EP849359-A2.
XX
PD 24-JUN-1998.
XX
PF 18-DEC-1997; 97EP-0122417.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
PI Oeda K, Wantanabe E;
XX
DR WPI; 1998-324670/29.

DR N-PSDB; AAV40801.
XX
PT New nucleic acid molecule encoding plant raffinose synthetase -
PT capable of producing raffinose, used as food additives with
PT beneficial effects on gastrointestinal flora
XX
PS Claim 1; Page 31-34; 44pp; English.
XX
CC This sequence represents the soybean raffinose synthetase of the
CC invention. The raffinose synthetase is capable of producing raffinose by
CC combining a D-galactosyl group through an alpha (1-6) bond with a
CC hydroxyl group attached to the carbon atom at position 6 of a D-glucose
CC residue in a sucrose molecule. The DNA can be used to modify metabolism
CC of a host organism by introducing into the host organism or its cell so
CC that the content of the raffinose family oligosaccharides in the host
CC organism or cell is changed. Raffinose oligosaccharides are useful as
CC food additives with beneficial effects on the gastrointestinal flora.
XX
SQ Sequence 781 AA:

Query Match 80.7%; Score 146; DB 19; Length 781;
Best Local Similarity 80.0%; Pred. No. 1.5e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 30
DB 220 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 249

RESULT 12
ID AAY30143 standard; Protein: 781 AA.
XX
AC AAY30143;
XX
DT 26-OCT-1999 (first entry)
XX
DE Amino acid sequence of a raffinose synthase protein.
XX
KW Raffinose synthase; plant; sucrose; raffinose.
XX
OS Glycine max.
XX
PN JP11215984-A.
XX
PD 10-AUG-1999.
XX
PF 12-DEC-1997; 97JP-0342899.
XX
PR 28-NOV-1997; 97JP-0329006.
XX
PR 18-DEC-1996; 96JP-0338673.
XX
PA (SUMO) SUMITOMO CHEM CO LTD.
XX
DR WPI; 1999-511112/43.
XX
DR N-PSDB; AA210002.
XX
PT New raffinose synthase gene - is prepared from a plant material
XX
PS Claim 8; Page 25-27; 40pp; Japanese.
XX
CC The present sequence represents a raffinose synthase protein. The
CC sequence is isolated from plant material. The protein forms raffinose
CC by complexing alpha(1 to 6)- D-galactosyl hydroxyl group of the 6C
CC of D-glucose residue in sucrose molecules.
XX
SQ Sequence 781 AA:

Query Match 80.7%; Score 146; DB 20; Length 781;
Best Local Similarity 80.0%; Pred. No. 1.5e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRHLVDGCG 30

```

Db      220 FGMCTWDAFYLVKHPGVEGVKGVLVEGSC   249
|||||
RESULT 13
AAB49400 AAB49400 standard; Protein; 781 AA.
AC AAB49400;
XX
DT 07-MAR-2001 (first entry)
XX
DE Soybean raffinose synthase.
XX
KW Plant promoter; transgenic plant; desired property.
XX
OS Glycine max.
XX
PN EP1048733-A2.
XX
PD 02-NOV-2000.
XX
PF 27-APR-2000; 2000EP-0108962.
XX
PR 30-APR-1999; 99JP-0124527.
PR 01-SEP-1999; 99JP-0247211.
XX
PA (SUMO ) SUMITOMO CHEM CO LTD.
XX
PI Ishige F, Watanabe E, Oeda K;
XX
WP1: 2001-104537/12.
DR N-PSDB; AAC89523.
XX
PT New soybean plant promoters useful for generating transgenic plants
XX With desired properties -
XX Example 6; Page 24-27; 36pp; English.
XX
PS The present invention provides novel plant promoters which can be used in
CC the production of transgenic plants which express genes with desired
CC properties.
XX
SQ Sequence 781 AA;

Query Match          80.7%; Score 146; DB 22; Length 781;
Best Local Similarity 80.0%; Pred.No.1.5e-12;
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0.

QY      1 FGMCTWDAFYLVKHPGVEGVGRHVDGSC   30
        |||||
Db      220 FGMCTWDAFYLVKHPGVEGVKGVLVEGSC   249

RESULT 14
ID AAY32075 AAY32075 standard; Protein; 572 AA.
AC AAY32075;
XX
DT 17-JAN-2000 (first entry)
XX
DE Rapeseed raffinose synthase.
XX
KW Raffinose synthase; rapeseed; transgenic plant.
XX
OS Brassica napus.
XX
FH Key Location/Qualifiers
FT Misc-difference 129 /note= "encoded by GGY"
FT Misc-difference 132 /note= "encoded by GCW"
FT
```

FT	Misc-difference	143	/note=	"encoded by CCS"
FT	Misc-difference	144	/note=	"encoded by tCR"
FT	Misc-difference	148	/note=	"encoded by CGR"
PN	EP953643-A2.			
XX	03-NOV-1999.			
XX	27-APR-1999;	99EP-0107430.		
XX	30-APR-1998;	98JP-0120550.		
XX	30-APR-1998;	98JP-0120551.		
XX	04-DEC-1998;	98JP-0345590.		
XX	10-DEC-1998;	98JP-0351246.		
XX	(SUMO) SUMITOMO CHEM CO LTD.			
XX	Watanabe E, Oeda K;			
XX	WPI: 1999-593144/51.			
XX	N-PSDB: AAZ20210.			
XX	New sense and antisense genes, useful for altering the level of			
XX	raffinose in food plants -			
XX	Claim 27: Page 36-38; 55pp: English.			
CC	This sequence represents rapessed raffinose synthase, a protein			
CC	that can bind a D-galactosyl group through an alpha(1-6) bond to the			
CC	hydroxy group attached to the carbon atom at the 6-position of the			
CC	D-glucose residue in a sucrose molecule to form raffinose. cDNA			
CC	(see AAZ20210) encoding the enzyme was isolated from rapessed cv.			
CC	Westar leaf cDNA by PCR. Probes or primers generated from plant			
CC	raffinose synthase genes (see AAZ20207-10) may be used to obtain			
CC	other raffinose synthase genes by labeled detection or amplification			
CC	(claimed). These genes may be used to control the levels of knock			
CC	raffinose produced in plants. Antisense genes can be used to knock			
CC	out existing gene activity, and sense genes to increase the level			
CC	of gene activity. The resulting transgenic plants may be used as a			
CC	food source to alter the growing conditions for gut enterobacteria,			
CC	providing general health advantages.			
XX	Sequence	572 AA;		
SO	Query Match	79.6%, Score 144, DB 20;	Length 572;	
	Best Local Similarity	80.0%; Pred. No. 2,1e-12;		
	Matches 24; Conservative 3; Mismatches 3;	Indels 0;	Gaps 0;	
OY	1 FGMCTWDAPFLTVHPGVIEGVHRHVPDGC 30			
Db	13 FGMCTWDAPFLTVHPDGVHKGVCCLVDGC 42			
RESULT 15				
ID	AAI32074			
XX	AAI32074 standard: Protein: 777 AA.			
XX	AAI32074;			
XX	17-JAN-2000 (first entry)			
XX	Mustard raffinose synthase.			
XX	Raffinose synthase; mustard; transgenic plant.			
XX	Brassica juncea.			
XX	Key Location/Qualifiers			
XX	Misc-difference 210	/note= "encoded by ACR"		

PX	
PN	EP953643-A2.
XX	
PD	03-NOV-1999.
XX	
PF	27-APR-1999; 99EP-0107430.
XX	
PR	30-APR-1998; 98JP-0120550.
PR	30-APR-1998; 98JP-0120551.
PR	04-DEC-1998; 98JP-0345590.
PR	10-DEC-1998; 98JP-0351246.
XX	
PA	(SUMO) SUMITOMO CHEM CO LTD.
XX	
PI	Watanabe E, Oeda K;
DR	WPI: 1999-593144/51.
XX	N-P.SDB; AAZ20209.
PT	New sense and antisense genes, useful for altering the level of
XX	raffinose in food plants -
PS	Claim 26: Page 29-31; 55pp: English.
XX	
CC	This sequence represents mustard raffinose synthase, a protein
CC	that can bind a D-galactosyl group through an alpha(1-6) bond to the
CC	hydroxy group attached to the carbon atom at the 6-position of the
CC	D-glucose residue in a sucrose molecule to form raffinose. CDNA
CC	(see AAZ20209) encoding the enzyme was isolated from mustard
CC	(Brassica juncea) leaf cDNA by PCR. Probes or primers generated from
CC	plant raffinose synthase genes (see AAZ20207-10) may be used to obtain
CC	other raffinose synthase genes by labeled detection or amplification
CC	(claimed). These genes may be used to control the levels of
CC	raffinose produced in plants. Antisense genes can be used to knock
CC	out existing gene activity, and sense genes to increase the level
CC	of gene activity. The resulting transgenic plants may be used as a
CC	food source to alter the growing conditions for gut enterobacteria,
CC	providing general health advantages.
XQ	Sequence 777 AA:

	Query Match	79.6%	Score 144	DB 20,	Length 777,
	Best Local Similarity	80.0%	Pred No. 3e-12;		
	Matches 24; Conservative	3;	Mismatches	3;	Indels 0;
			Gaps	0;	
Oy	1 FGCWDAFYLTVMFGVIEGYRHLVDGCG	30			
Db	218 FGCWDAFYLTVMFDGVHKKVCCLVDGGC	247			

Search completed: March 27, 2003, 10:03:41
Job time : 22.2195 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 8.59756 Seconds
(without alignments)
335.448 Million cell updates/sec

Title: US-09-675-208-1
Perfect score: 181
Sequence: 1 FGWCTWDAFYLTVPQGVIEGVRLVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR_73:*
2: PIR1:*
3: PIR2:*
4: PIR3:*
5: PIR4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	61.3	765	2 S45033	probable imbibition
2	108	59.7	357	2 T09530	probable seed inh1
3	108	59.7	757	2 S27762	S1p1 protein - bar
4	108	59.7	773	2 T46188	Imbibition protein
5	106	58.6	807	2 C85025	hypothetical prote
6	82	45.3	1170	2 C96599	protein F20N2.14 l
7	72	39.8	371	2 T01717	hypothetical prote
8	64.5	35.6	1374	2 E82168	probable alpha-1,6
9	60	33.1	137	2 I46711	fibroblast growth
10	60	33.1	146	1 S00185	basic fibroblast g
11	60	33.1	154	2 A31674	basic fibroblast g
12	60	33.1	154	2 C37360	basic fibroblast g
13	60	33.1	157	1 GKB0B	basic fibroblast g
14	60	33.1	210	2 A32398	basic fibroblast g
15	60	33.1	451	2 T24018	hypothetical prote
16	59	32.6	164	2 S31622	basic fibroblast g
17	57.5	31.8	649	2 D90496	hypothetical prote
18	56.5	31.2	373	2 A10440	probable aliphatic
19	53.5	29.6	319	2 C90756	hypothetical prote
20	53.5	29.6	333	2 A85620	hypothetical prote
21	53.5	29.6	333	2 G64833	probable nitrate t
22	53	29.3	189	2 A48834	basic fibroblast g
23	53	29.3	557	2 E83744	hypothetical prote
24	52.5	29.0	265	2 E82964	membrane subunit o
25	52	28.7	237	2 B86420	unknown protein, 1
26	52	28.7	619	2 D86509	S/T protein kinase
27	52	28.7	619	2 D81556	conserved hypothet
28	52	28.7	619	2 A72114	s/t protein kinase
29	52	28.7	1462	1 B36182	protein-tyrosine-p

30	51.5	28.5	353	2 AE2067	histidinol phosph
31	51	28.2	106	1 ERAD31	early E3A 12.1K pr
32	51	28.2	106	2 S52799	hypothetical prote
33	51	28.2	106	2 S52809	hypothetical prote
34	51	28.2	225	2 AD0798	probable phosphat
35	51	28.2	1039	2 F71427	hypothetical prote
36	50	27.6	188	2 AD0533	conserved hypothet
37	50	27.6	284	2 AB0521	AMPE protein (limp
38	50	27.6	381	2 T23250	hypothetical prote
39	50	27.6	996	2 F98092	cytM protein, cyto
40	50	27.6	996	2 A85228	bacteriocin format
41	49.5	27.3	243	2 AC3600	cellulase (EC 3.2.
42	49.5	27.3	310	2 S39592	prolyl aminopeptid
43	49.5	27.3	720	2 AF0242	probable exported
44	49	27.1	79	2 S55456	GAM1 protein - Pla
45	49	27.1	191	1 H64744	yaed protein - Esc

ALIGNMENTS

RESULT 1

S45033
Probable imbibition protein - wild cabbage
C:Species: Brassica oleracea (wild cabbage)
C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Sep-1997
C:Accession: S45033
R:Fujikura, Y.; Karsen, C.K.
Submitted to the EMBL Data Library, May 1994
A:Description: Cauliflower cDNA encoding a putative imbibition protein.
A:Reference number: S45033
A:Accession: S45033
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-765 <FNU>
A:Cross-references: EMBL:X79330; NID:q488786; PID:q488787

Query Match 61.3%; Score 111; DB 2; Length 765;
Best Local Similarity 65.5%; Pred. No. 2.3e-07;
Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRLVDGCG 29

DB 198 FGWCTWDAFYLTVPQGVIEGVRLVDGCG 226

RESULT 2

T09530
Probable seed imbibition protein - chickpea (fragment)
C:Species: Cicer arietinum (chickpea, garbanzo)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 20-Jun-2000
C:Accession: T09530
R:Cervantes, E.
Submitted to the EMBL Data Library, February 1996
A:Reference number: Z16718
A:Accession: T09530
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-357 <CEP>
A:Cross-references: EMBL:X95875; NID:q1212811
A:Experimental source: germinating seed
C:Genetics:
A:Gene: s1p

Query Match 59.7%; Score 108; DB 2; Length 357;
Best Local Similarity 62.1%; Pred. No. 2.6e-07;
Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 FGWCTWDAFYLTVPQGVIEGVRLVDGCG 29

DB 31 FGWCTWDAFYLTVPQGVIEGVRLVDGCG 59

```
RESULT 3
S27762
S1p1 Protein - barley
C:Species: Hordeum vulgare (barley)
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 09-Sep-1997
C:Accession: S27762
R:Heck, G.R.; Dorsett, C.; Ho, T.H.
submitted to the EMBL Data Library, February 1992
A:Description: Cloning and characterization of a gene, S1p1, associated with seed imbibition
A:Reference number: S27762
A:Accession: S27762
A:Molecule type: DNA
A:Residues: 1-757 <HEC>
A:Cross-references: EMBL:M7475; NID:g167099; PID:g167100
C:Genetics:
A:introns: 64/2; 151/1; 181/2; 197/3; 489/3; 557/2

Query Match
Best Local Similarity 59.7%; Score 108; DB 2; Length 757;
Matches 18; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLTVPQGVIEGVRHLVDG 29
||||||| | | | | | | | |
Db 205 FGMCTWDAFYLTVDATGKQGLRLAEGG 233

RESULT 4
T46188
Imbibition protein homolog - Arabidopsis thaliana
N:Alternate names: protein T8H10.120
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000
C:Accession: T46188
R:Benes, V.; Rechmann, S.; Borkova, D.; Ansoorge, W.; Mewes, H.W.; Lemcke, K.; Mayer, K.E.
submitted to the Protein Sequence Database, January 2000
A:Reference number: 223014
A:Accession: T46188
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-773 <BEN>
A:Cross-references: EMBL:ALJ3248
A:Experimental source: cultivar Columbia; BAC clone T8H10
C:Genetics:
A:Map position: 3
A:introns: 64/2; 146/1; 176/2; 192/3; 223/2; 259/2; 300/2; 484/3; 507/2; 552/2; 625/1; 6
A:Note: T8H10.120

Query Match
Best Local Similarity 59.7%; Score 108; DB 2; Length 773;
Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLTVPQGVIEGVRHLVDG 29
||||||| | | | | | | | |
Db 200 FGMCTWDAFYLTVDATGKQGLRLAEGG 228

RESULT 5
C85025
Hypothetical protein AT4g01970 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: C85025
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488; PMID:10617198
A:Accession: C85025
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-807 <STO>
A:Cross-references: GB:NC_001268; NID:g7268581; PIDN:CAB80690.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g01970
```

```
A:Map position: 4

Query Match
Best Local Similarity 58.6%; Score 106; DB 2; Length 807;
Matches 18; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLTVPQGVIEGVRHLVDG 29
||||||| | | | | | | | |
Db 235 FGMCTWDAFYLTVDPATMTGVKEFEDGG 263

RESULT 6
C96599
protein F20N2.14 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: C96599
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Matzila,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: C96599
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1170 <STO>
A:Cross-references: GB:AE005173; NID:g8778496; PIDN:AAF79504.1; GSPDB:GN00141
C:Genetics:
A:Gene: F20N2.14
A:Map position: 1

Query Match
Best Local Similarity 45.3%; Score 82; DB 2; Length 1170;
Matches 14; Conservative 3; Mismatches 10; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLTVPQGVIEGVRHLVD 27
||||||| | | | | | | | |
Db 203 FGMCTWDAFYLTNTAKDVKGGLSNCND 229

RESULT 7
T01717
Hypothetical protein A_I5002N01.5 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01717
R:Scheef, P.; Magft, L.
submitted to the EMBL Data Library, June 1997
A:Description: The sequence of A. thaliana I5002N01.
A:Reference number: 214407
A:Accession: T01717
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-371 <SCH>
A:Cross-references: EMBL:AF007269; NID:g2191126; PID:g2191156
A:Experimental source: cultivar Columbia
C:Genetics:
A:Map position: 4
A:introns: 5/2; 25/3; 103/3; 126/1; 174/2; 240/1; 284/1; 295/1; 333/2; 349/2
A:Note: A_I5002N01.5

Query Match
Best Local Similarity 39.8%; Score 72; DB 2; Length 371;
Matches 11; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 FGMCTWDAFYLTVP 13
||||||| |
```

Db 111 FGCTWDAFYRKV 123

RESULT 8

Probable alpha-1,6-galactosidase VC1690 [Imported] - Vibrio cholerae (strain N16961 serc

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: E82168

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
L, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:10952301

A:Accession: E82168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-374 <HEI>

A:Cross-references: GB:AE004247; GB:AE003852; NID:g956204; PIDN:AAE94840.1; GSPDB:GN001

C:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC1690

A:Map position: 1

Query Match 35.6%; Score 64.5; DB 2; Length 374;

Best Local Similarity 38.5%; Pred. No. 0.32;

Matches 15; Conservative 4; Mismatches 9; Indels 11; Gaps 2;

Oy 2 GCCTWDAFYLVHPOGVIEGV-----RH-----LVDCG 29

Db 197 GWCWYAYAEVTEODIKENVAIIAERHPELEWVLLDDG 235

RESULT 9

146711 fibroblast growth factor - rabbit (fragment)

C:Species: Oryctolagus cuniculus (domestic rabbit)

C:Date: 14-Feb-1997 #sequence_revision 14-Feb-1997 #text_change 16-Jul-1999

C:Accession: I46711

R:Winkler, J.A.; Erlens, R.; Alberts, G.F.; Janat, M.F.; Lian, G.

Ann. N. Y. Acad. Sci. 143, 518-527, 1993

A:Title: Elevated expression of basic fibroblast growth factor in an immortalized rabbit

A:Reference number: I46711; MUID:93343209; PMID:8342599

A:Accession: I46711

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-137 <WIN>

A:Cross-references: GB:L12034; NID:g165014; PIDN:AAA31248.1; PID:g165015

C:Superfamily: fibroblast growth factor

Query Match 33.1%; Score 60; DB 2; Length 137;

Best Local Similarity 40.0%; Pred. No. 0.47;

Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTWDAFYLVHPOGVIEGRHLVD 27

Db 24 YCKNGFFLRHHPDGRVDGVRKSD 48

RESULT 10

S00185

basic fibroblast growth factor - sheep

N:Alternate names: prostathropin

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

C:Accession: S00185

R:Simpson, R.J.; Moritz, R.L.; Lloyd, C.J.; Fabri, L.J.; Nice, E.C.; Rudira, M.R.; Burge

FEBS Lett. 224, 128-132, 1987

A:Title: Primary structure of ovine pituitary basic fibroblast growth factor.

A:Reference number: S00185; MUID:88055577; PMID:3678486

A:Accession: S00185

A:Molecule type: protein

A:Residues: 1-146 <SIM>

C:Superfamily: fibroblast growth factor

C:Keywords: growth factor; heparin binding; mitogen

F:18-22/Region: heparin binding #status predicted

F:107-110/Region: heparin binding #status predicted

Query Match 33.1%; Score 60; DB 1; Length 146;

Best Local Similarity 40.0%; Pred. No. 0.5;

Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTWDAFYLVHPOGVIEGRHLVD 27

Db 24 YCKNGFFLRHHPDGRVDGVRKSD 48

RESULT 11

A31674 basic fibroblast growth factor precursor - rat

N:Alternate names: bFGF

C:Species: Rattus norvegicus (Norway rat)

C:Date: 21-May-1990 #sequence_revision 21-May-1990 #text_change 16-Jul-1999

C:Accession: A31674; S00876; S24309

R:Shimasaki, S.; Emoto, N.; Koba, A.; Mercado, M.; Shibata, F.; Cooksey, K.; Baird, A

Biochem. Biophys. Res. Commun. 157, 236-263, 1988

A:Title: Complementary DNA cloning and sequencing of rat ovarian basic fibroblast gro

A:Reference number: A31674; MUID:89061721; PMID:3196337

A:Accession: A31674

A:Molecule type: mRNA

A:Residues: 1-154 <SHI>

A:Cross-references: GB:M22427; NID:g204285; PIDN:AAA41210.1; PID:g204286

R:Kurokawa, T.; Seno, M.; Igarashi, K.

Nucleic Acids Res. 16, 5201, 1988

A:Title: Nucleotide sequence of rat basic fibroblast growth factor cDNA.

A:Reference number: S00876; MUID:88262516; PMID:3387229

A:Accession: S00876

A:Molecule type: mRNA

A:Residues: 1-154 <KUR>

A:Cross-references: EMBL:X07285; NID:g56203; PIDN:CAA30265.1; PID:g56204

R:El-Husseni, A.E.D.; Patterson, J.A.; Myal, Y.; Shiu, R.P.C.

Biochim. Biophys. Acta 1131, 314-316, 1992

A:Title: PCR detection of the rat brain basic fibroblast growth factor (bFGF) mRNA co

A:Reference number: S24309; MUID:92329546; PMID:1378302

A:Accession: S24309

A:Status: preliminary; translation not shown

A:Molecule type: mRNA

A:Residues: 35-154 <ELH>

A:Cross-references: EMBL:X61697; NID:g56143; PIDN:CAA43863.1; PID:g56144

C:Superfamily: fibroblast growth factor

C:Keywords: growth factor

F:1-9/Domain: signal sequence #status predicted <SIG>

F:10-154/Product: basic fibroblast growth factor #status predicted <MAT>

Query Match 33.1%; Score 60; DB 2; Length 154;

Best Local Similarity 40.0%; Pred. No. 0.53;

Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTWDAFYLVHPOGVIEGRHLVD 27

Db 32 YCKNGFFLRHHPDGRVDGVRKSD 56

RESULT 12

C37360

basic fibroblast growth factor - mouse

C:Species: Mus musculus (house mouse)

C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 16-Jul-1999

C:Accession: C37360

R:Heberly, J.M.; Basilico, C.; Goldfarb, M.; Haub, O.; Martin, G.R.

Dev. Biol. 138, 454-463, 1990

A:Title: Isolation of cDNAs encoding four mouse FGF family members and characterizati

A:Reference number: A37360; MUID:90201563; PMID:2318343

A:Accession: C37360

A:Status: preliminary

Growth Factor 4 277-287, 1991
A:Title: Functional characterization of the human basic fibroblast growth factor gene
A:Reference number: A61537; MUID:92110035; PMID:1764264
A:Accession: A61537
A:Molecule type: DNA
A:Residues: 1-114 <SH2>
A:Note: authors translated the codon GGA for residue 47 as Ala
R:Kurokawa, T.; Sasada, R.; Iwane, M.; Igatahshi, K.
FEBS Lett. 213, 189-194, 1987
A:Title: Cloning and expression of cDNA encoding human basic fibroblast growth factor.
A:Reference number: A26642; MUID:87162468; PMID:2435575
A:Accession: A26642
A:Molecule type: mRNA
A:Residues: 56-210 <KUR>
A:Cross-references: GB:M77968; NID:g182562; PIDN:AAA5248.1; PID:g182563
R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergle, A.; Fiddes, J.C.
Cold Spring Harb. Symp. Quant. Biol. 51, 657-668, 1986
A:Title: Human basic fibroblast growth factor: nucleotide sequence, genomic organization
A:Reference number: A90924; MUID:87211066; PMID:3472745
A:Accession: B32878
A:Molecule type: mRNA
A:Residues: 56-210 <ABR>
A:Note: the authors translated the codon GAA for residue 108 as Gly
R:Abraham, J.A.; Whang, J.L.; Tumolo, A.; Mergle, A.; Friedman, J.; Gospodarowicz, D.; E
EMBO J. 5, 2523-2528, 1986
A:Title: Human basic fibroblast growth factor: nucleotide sequence and genomic organization
A:Reference number: S00297; MUID:87053817; PMID:3780670
A:Accession: S00297
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-155 <AB2>
A:Note: the authors translated the codon GAA for residue 108 as Gly
R:Shimoyama, Y.; Gotou, M.; Ino, Y.; Sakamoto, M.; Kato, K.; Hirohashi, S.
Jpn. J. Cancer Res. 82, 1263-1270, 1991
A:Title: Characterization of high-molecular-mass forms of basic fibroblast growth factor
in rat gingiva.
A:Reference number: A54316; MUID:92091228; PMID:1772615
A:Accession: A54316
A:Molecule type: protein
A:Residues: 'xx', '86-88', 'x', '90-91', 'x', '93-95 <SH3>
A:Experimental source: C-1121 hepatocellular carcinoma cell line
A:Note: sequence extracted from NCBI backbone (NCBIP:71595)
A:Accession: B54316
A:Molecule type: protein
A:Residues: 'xxx', '19', 'x', '21-29 <SH2>
A:Note: sequence extracted from NCBI backbone (NCBIP:71594)
R:Fejge, J.J.; Bradley, J.D.; Fryburg, K.; Farris, J.; Cousens, L.C.; Barr, P.J.; Baird
J. Cell Biol. 109, 3105-3114, 1989
A:Title: Differential effects of heparin, fibronectin, and laminin on the phosphorylation
A:Reference number: A53624; MUID:90078343; PMID:2392418
A:Accession: A53624
A:Status: preliminary
A:Molecule type: protein
A:Residues: 57-210 <FEI>
A:Status: preliminary
A:Molecule type: protein
R:Story, M.T.; Esch, F.; Shimaski, S.; Sasse, J.; Jacobs, S.C.; Lawson, R.K.
Biochem. Biophys. Res. Commun. 142, 702-709, 1987
A:Title: Amino-terminal sequence of a large form of basic fibroblast growth factor isolate
A:Reference number: A25824; MUID:87156686; PMID:2435284
A:Accession: A25824
A:Molecule type: protein
A:Residues: 57-77 <STO>
A:Experimental source: prostate
R:Giemenes-Gallego, G.; Conn, G.; Hatcher, V.B.; Thomas, K.A.
Biochem. Biophys. Res. Commun. 135, 541-548, 1986
A:Title: Human brain-derived acidic and basic fibroblast growth factors: amino terminal
A:Reference number: A90122; MUID:86186784; PMID:3964259
A:Accession: B24243
A:Molecule type: protein
A:Residues: 65-102, 'x', '104-105 <GIM>
A:Experimental source: brain
R:Gautschi, P.; Frazer-Schroder, M.; Bohlen, P.
FEBS Lett. 204, 203-207, 1986
A:Title: Partial molecular characterization of endothelial cell mitogens from human brain

A:Reference number: A91364; MUID:86275260; PMID:3722516
A:Molecule type: Protein
A:Residues: 65-88,'X','90-98','X',100 <GAV>
R:Sommer,A.; Brewer,M.T.; Thompson,R.C.; Moscattelli,D.; Presta,M.; Rifkin,D.B.
Biochem. Biophys. Res. Commun. 144, 543-550, 1987
A>Title: A form of human basic fibroblast growth factor with an extended amino termin
A:Reference number: S42242; MUID:87213238; PMID:3579930
A:Accession: S42242
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 54-210 <SOM>
A:Cross-references: EMBL:M17599; NID:g183086; PIDN:AA52534.1; PID:g183087
R:Pantoliaro,M.W.; Horlick,R.A.; Springer,B.A.; Van Dyk,D.E.; Toberty,T.; Wetmore
Biochemistry 33, 10229-10248, 1994
A>Title: Multivalent ligand-receptor binding interactions in the fibroblast growth fa
A:Reference number: A55784; MUID:94347757; PMID:7520751
A:Accession: B55784
A:Molecule type: protein
A:Residues: 54-71 <PAN>
R:Watson,R.; Anthony,F.; Pickett,M.; Lambden,P.; Masson,G.M.; Thomas,E.J.
Biochem. Biophys. Res. Commun. 187, 1227-1231, 1992
A>Title: Reverse transcription with nested polymerase chain reaction shows expres
lients.
A:Reference number: I52267; MUID:93038590; PMID:1417798
A:Accession: I52267
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 95-182 <RES>
A:Cross-references: GB:S47380; NID:g256535; PIDN:AAD13853.1; PID:g4261553
A:Experimental source: granulosa cells
R:Patry,V.; Bugler,B.; Amarlic,F.; Promé,J.C.; Prats,H.
FEBS Lett. 349, 23-28, 1994
A>Title: Purification and characterization of the 210-amino acid recombinant basic fi
A:Reference number: S46253; MUID:94320639; PMID:8045296
A:Accession: S46253
A:Molecule type: protein
A:Residues: 39-53;65-88 <PAT>
A>Note: recombinant gene expressed in Escherichia coli
C:Genetics:
A:Gene: GDB:F2GF2; FGFB
A:Cross-references: GDB:I19910; OMIM:134920
A:Map position: Aq25-q27
A:Start codon: CTG
C:Superfamily: fibroblast growth factor
C:Keywords: alternative initiators; angiogenesis; growth factor; heparin binding; multi
F.1-210/Product: basic fibroblast growth factor, 22-5K form #status predicted <MA2>
F.65-210/Product: basic fibroblast growth factor, 18K form #status predicted <MA7>
F.82-86/Region: heparin binding #status predicted
F.I71-I74/Region: heparin binding #status predicted

Query Match 33.1%; Score 60; DB 2; Length 210;
Best Local Similarity 40.0%; Pred. NO. 0.73;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTWDAEYLTVHPQGVIEGRHLVD 27
.: : : : : : : : : :
Db 88 YCKNGGEFLRIHPDGRVDCVRKSD 112

RESULT 15
T24018
hypothetical protein R0787.11 - Caenorhabditis elegans
CSpecies: Caenorhabditis elegans
C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
C.Accession: T24018
R:Harris,B.
submitted to the EMBL Data Library, July 1996
A:Reference number: Z19830
A:Accession: T24018
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-451 <WII>

A:Cross-references: EMBL:Z75955; PIDN:CAB00120.1; GSPDB:GN00023; CESP:R07B7.11
 A:Experimental source: clone R07B7
 C:Genetics:
 A:Gene: CESP:R07B7.11
 A:Map position: 5
 A:Introns: 61/1, 107/3, 317/3; 351/3
 C:Superfamily: alpha-galactosidase

Query Match 33.1%; Score 60; DB 2; Length 451;
 Best Local Similarity 36.8%; Pred. No. 1.7;
 Matches 14; Conservative 3; Mismatches 11; Indels 10; Gaps 2;

Qy 2 GWCCTDPAFLTV---HPQGV-----EGVRHLVDG 29
 ||:|||||:|||||:|||||
 Db 27 GWMSTAFYCEIDCVKHPGTCINEQLYKDMADQIVSGG 64

Search completed: March 27, 2003, 10:06:47
 Job time : 10.5976 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:01:35 ; Search time 4.93902 Seconds
(Without alignments)
251.930 Million cell updates/sec

Title: US-09-675-208-1

Perfect score: 181
Sequence: 1 FGWCTMDAFYLTWHPQGVIEGVRLVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	60	33.1	137	1	FGF2_RABIT
2	60	33.1	154	1	FGF2_MOUSE
3	60	33.1	154	1	FGF2_RAT
4	60	33.1	155	1	FGF2_BOVIN
5	60	33.1	155	1	FGF2_HUMAN
6	60	33.1	155	1	FGF2_SHEEP
7	59	32.6	156	1	FGF2_MONDO
8	53.5	29.6	319	1	SSUA_ECOLI
9	53	29.3	158	1	FGF2_CHICK
10	52	28.7	1462	1	PRP6_DROME
11	51.5	28.5	353	1	H181_AMASP
12	51	28.2	106	1	E312_ADE03
13	49.5	27.3	310	1	PIP_NEIGO
14	49	27.1	191	1	VAED_ECOLI
15	49	27.1	227	1	LIPIB_STY3
16	49	27.1	390	1	DXR_FUSNN
17	48.5	26.8	302	1	YIAL_PSESS
18	48.5	26.8	515	1	GLCM_MOUSE
19	48.5	26.8	536	1	GLCM_HUMAN
20	48	26.5	207	1	BSFG_HUMAN
21	48	26.5	207	1	BSFG_RAT
22	48	26.5	366	1	CHSB_IPORF
23	48	26.5	384	1	DXR_CLOPE
24	48	26.5	384	1	DXR_ZYMO
25	48	26.5	468	1	LACG_STRMU
26	48	26.5	470	1	LACG_STANU
27	47.5	26.2	369	1	MAGA_HUMAN
28	47.5	26.2	500	1	C72X_ARATH
29	47.5	26.2	500	1	C72Y_ARATH
30	47	26.0	247	1	FGFE_HUMAN
31	47	26.0	247	1	FGFE_MOUSE
32	47	26.0	341	1	SP12_COMPX
33	47	26.0	361	1	U204_ARATH

34	47	26.0	382	1	DXR_THETN	O8a28 thermomae
35	47	26.0	406	1	DXR_MYCLE	O9cbu3 mycobacteri
36	47	26.0	413	1	DXR_MYCTU	O10798 mycobacteri
37	47	26.0	548	1	HEML_YEAST	P09950 saccharomyc
38	47	26.0	686	1	VILL_HUMAN	O15195 homo sapien
39	46.5	25.7	299	1	RFBJ_SALT	P22716 salmonella
40	46.5	25.7	498	1	NDDO_ALCXH	P94212 alcaligenes
41	46.5	25.7	521	1	OPN4_MOUSE	O9qxx9 mus musculu
42	46.5	25.7	582	1	HEMO_OPSTA	P43090 opsanus lau
43	46	25.4	132	1	VEA_HPVA	P25483 human papil
44	46	25.4	222	1	SP1B_VACC	P20842 vaccinia vi
45	46	25.4	352	1	CHSC_IPOPU	P48399 ipomoea pur

ALIGNMENTS

RESULT 1	ID	FGF2_RABIT	STANDARD:	PRT:	137 AA.
AC	P48799	01-FEB-1996 (Rel. 33, Created)			
DT	01-FEB-1996 (Rel. 33, Last sequence update)				
DT	15-JUN-2002 (Rel. 41, Last annotation update)				
DE	Heparin-binding growth factor 2 (HBGF-2) (Basic fibroblast growth factor) (BFGF) (Prostathelin) (Fragment).				
GN	FGF2.				
OS	Oryctolagus cuniculus (Rabbit).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.				
OX	NCBI_TaxID=9986;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=New Zealand white; TISSUE=Smooth muscle;				
RX	MEDLINE=93343209; PubMed=8342599;				
RA	Winkles J.A., Friesel R., Alberts G.F., Janat M.F., Liu G.;				
RT	"Elevated expression of basic fibroblast growth factor in an				
RL	Am. J. Pathol. 143:518-527(1993).				
CC	- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS				
CC	IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN				
CC	VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND				
CC	CONCENTRATION OF THESE 2 GROWTH FACTORS.				
CC	- SUBUNIT: MONOMER.				
CC	- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES				
CC	ACEF.				
CC	- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: L12034; AAA31248.1; -				
DR	HSSP: P09038; 1BFF.				
DR	InterPro: IPR002209; HB/F-growthfact.				
DR	Pfam: PF00167; FGF, 1.				
DR	ProDom: PD000831; HB/F-growthfact; 1.				
DR	SMART: SM00442; FGF, 1.				
DR	PROSITE: PS00247; HBGF_FGF, 1.				
KW	Growth factor; Mitogen; Angiogenesis; Heparin-binding.				
FT	BINDING 18 22 HEPARIN (POTENTIAL).				
FT	BINDING 107 110 HEPARIN (POTENTIAL).				
FT	NON_TER 137 137				
SQ	SEQUENCE 137 AA; 15418 MW; 0D9EE457B88B8C51 CRC64;				
Query Match	33.1%; Score 60; DB 1; Length 137;				
Best Local Similarity	40.0%; Pred. No. 0.18;				
Matches	10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;				

DR PIR: S00876; S00876.
DR PIR: A31674; A31674.
DR HSSP: P09038; 1BFE.
DR InterPro: IPR002309; HB/F_growthfact.
DR InterPro: IPR002348; IL1_HBGF.
DR Pfam: PF00167; FGF; 1.
DR PRINTS: PR00262; IL1HBGF.
DR ProDom: PD000831; HB/F_growthfact; 1.
DR SMART: SM00442; FGF; 1.
DR PROSITE: PS00247; HBGF_FGF; 1.
KM Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 9
FT CHAIN 10 154 HEPARIN-BINDING GROWTH FACTOR 2.
FT BINDING 26 30 HEPARIN (POTENTIAL).
FT BINDING 115 118 HEPARIN (POTENTIAL).
SQ SEQUENCE 154 AA: 17139 MW: 1A0F14FA23D8403 CRC64;

Query Match 33.1%; Score 60; DB 1; Length 154;
Best Local Similarity 40.0%; Pred. No. 0.21;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

QY 3 WCTMDAFYTVHPOGVIEGVRLVLD 27
DB 32 YCKNGCFELRHHPDGRVDSVREKSD 56

RESULT 4
FGF2_BOVIN STANDARD: PRT: 155 AA.
ID FGF2_BOVIN STANDARD: PRT: 155 AA.
AC P03969;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (basic fibroblast growth factor) (BFGF) (Prostatropin) [contains: kidney-derived growth factor].
DE FGF2 OR FGF-2.
GN Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; CC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-86261806; PubMed-2425435;
RA Abraham J.A., Mergia A., Whang J.L., Tumolo A., Friedman J., Hjerlild K.A., Gospodarowicz D., Fiddes J.C.;
RT "Nucleotide sequence of a bovine clone encoding the angiogenic protein, basic fibroblast growth factor.";
RL Science 233:545-548(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-87217066; PubMed-3472745;
RA Abraham J.A., Whang J.L., Tumolo A., Mergia A., Fiddes J.C.;
RT "Human basic fibroblast growth factor: nucleotide sequence, genomic organization, and expression in mammalian cells.";
RL Cold Spring Harb. Symp. Quant. Biol. 51:657-668(1986).
RN [3]
RP SEQUENCE OF 10-155.
RX MEDLINE-86016731; PubMed-3863109;
RA Esch F., Baird A., Ling N., Ueno N., Hill F., Denotroy L., Klepper R., Gospodarowicz D., Boehlen P., Guillemin R.;
RT "Primary structure of bovine pituitary basic fibroblast growth factor (FGF) and comparison with the amino-terminal sequence of bovine brain acidic FGF.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6507-6511(1985).
RN [4]
RP SEQUENCE OF 1-9.
RX MEDLINE-86295737; PubMed-3741423;
RA Ueno N., Baird A., Esch F., Ling N., Guillemin R.;
RT "Isolation of an amino terminal extended form of basic fibroblast growth factor.";
RL Biochem. Biophys. Res. Commun. 138:580-588(1986).

RN [5]
RP SEQUENCE OF 25-41.
RC TISSUE=Kidney;
RX MEDLINE-86095426; PubMed-4081126;
RA Baird A., Esch F., Boehlen P., Ling N., Gospodarowicz D.;
RT "Isolation and partial characterization of an endothelial cell growth factor from the bovine kidney: homology with basic fibroblast growth factor.";
RL Regul. Pept. 12:201-213(1985).
RN [6]
RP SEQUENCE OF 21-40.
RC TISSUE=Kidney;
RX MEDLINE-87119165; PubMed-3809608;
RA Ueno N., Baird A., Esch F., Shimasaki S., Ling N., Guillemin R.;
RT "Purification and partial characterization of a mitogenic factor from bovine liver: structural homology with basic fibroblast growth factor.";
RL Regul. Pept. 16:135-145(1986).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (3.0 ANGSTROMS).
RX MEDLINE-91095983; PubMed-1702556;
RA Zhu X., Komiyama H., Chirino A., Faham S., Fox G.M., Arakawa T., Hsu B.T., Rees D.C.;
RT "Three-dimensional structures of acidic and basic fibroblast growth factors.";
RL Science 251:90-93(1991).
CC -!- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC -!- SUBUNIT: MONOMER.
CC -!- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES AFGF.
CC -!- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M13440; AAA30518.1; -;
DR PIR: A24663; GKBOB.
DR PIR: A24819; A24819.
DR PIR: A32878; A32878.
DR PDB: 1BAS; 3I-OCT-93.
DR InterPro: IPR002209; HB/F_growthfact.
DR InterPro: IPR002348; IL1_HBGF.
DR Pfam: PF00167; FGF; 1.
DR PRINTS: PR00262; IL1HBGF.
DR ProDom: PD000831; HB/F_growthfact; 1.
DR SMART: SM00442; FGF; 1.
DR PROSITE: PS00247; HBGF_FGF; 1.
KM Growth factor; Mitogen; Angiogenesis; Heparin-binding;
RN 3D-structure.
FT PROPEP 1 9
FT CHAIN 10 155 HEPARIN-BINDING GROWTH FACTOR 2.
FT SITE 25 155 KIDNEY-DERIVED GROWTH FACTOR.
FT SITE 46 48 CELL ATTACHMENT SITE (POTENTIAL).
FT SITE 88 90 CELL ATTACHMENT SITE (POTENTIAL).
FT BINDING 27 31 HEPARIN (POTENTIAL).
FT BINDING 116 119 HEPARIN (POTENTIAL).
FT STRAND 30 34
FT TURN 35 38
FT STRAND 39 43
FT TURN 44 46
FT STRAND 49 52
FT TURN 53 56
FT HELIX 58 60
FT STRAND 62 68
FT TURN 69 70

[illegible]

[5]
SEQUENCE FROM N.A.
MEDLINE=89184522; PubMed=2538817;
Ra Priats H., Kaghani M., Prats A.C., Klagsbrun M., Lelias J.M.,
Ra Liazun P., Chalou N., Tauber J., Amalric F., Smith J.A.,
Ra Caput D.;
"High molecular mass forms of basic fibroblast growth factor are
initiated by alternative CUG codons";
Proc. Natl. Acad. Sci. U.S.A. 86:1836-1840(1989).
[6]
SEQUENCE OF 10-35.
MEDLINE=86275260; PubMed=3732516;
Ra Gautschi P., Frater-Schroeder M., Boehlen P.;
"Partial molecular characterization of endothelial cell mitogens from
human brain: acidic and basic fibroblast growth factors";
FEBS Lett. 204:203-207(1986).
[7]
SEQUENCE OF 16-39.
MEDLINE=86186784; PubMed=3964255;
Ra Gluener-Galligo G., Conn G., Hatcher V.B., Thomas K.A.;
"Human brain-derived acidic and basic fibroblast growth factors:
amino terminal sequences and specific mitogenic activities";
Biochem. Biophys. Res. Commun. 135:541-548(1986).
[8]
SEQUENCE OF 2-22.
MEDLINE=87156686; PubMed=2433284;
Ra Story M.T., Esch F., Shimazaki S., Sasse J., Jacobs S.C., Lawson R.K.,
Ra "Amino-terminal sequence of a large form of basic fibroblast growth
factor isolated from human benign prostatic hyperplastic tissue.";
Biochem. Biophys. Res. Commun. 142:702-709(1987).
[9]
X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
MEDLINE=911195367; PubMed=1707542;
Ra Eriksson A.E., Couzens L.S., Weaver L.H., Matthews B.W.;
"Three-dimensional structure of human basic fibroblast growth
factor";
Proc. Natl. Acad. Sci. U.S.A. 88:3441-3445(1991).
[10]
X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
MEDLINE=94004464; PubMed=7691311;
Ra Eriksson A.E., Couzens L.S., Matthews B.W.;
"Refinement of the structure of human basic fibroblast growth factor
at 1.6-A resolution and analysis of presumed heparin binding sites by
selenate substitution";
Protein Sci. 2:1274-1284(1993).
[11]
X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
MEDLINE=91195368; PubMed=1849658;
Ra Zhang J., Couzens L.S., Barr P.J., Sprang S.R.;
"Three-dimensional structure of human basic fibroblast growth factor,
a structural homolog of interleukin 1 beta.";
Proc. Natl. Acad. Sci. U.S.A. 88:3446-3451(1991).
[12]
X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).
MEDLINE=92121151; PubMed=1769964;
Ra Ago H., Kitaawa Y., Fujishima A., Matsunura Y., Katsube Y.;
"Crystal structure of basic fibroblast growth factor at 1.6-A
resolution";
R. Biochem. 110:360-363(1991).
[13]
X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS).
MEDLINE=91095983; PubMed=1702556;
Ra Zhu X., Komiya H., Chirino A., Faham S., Fox G.M., Arakawa T.,
Ra Hsu B.T., Rees D.C.;
"Three-dimensional structures of acidic and basic fibroblast growth
factors.";
Science 251:90-93(1991).
[14]
STRUCTURE BY NMR.
MEDLINE=97040521; PubMed=8885834;
Ra Moy F.J., Seddon A.P., Boehlen P., Powers R.;
"High-resolution solution structure of basic fibroblast growth factor
determined by multidimensional heteronuclear magnetic resonance

```
CC FT STRAND      103    107
CC FT TURN       109    110
CC FT STRAND     113    117
CC FT TURN       121    122
CC FT STRAND     124    124
CC FT STRAND     127    127
CC FT TURN       129    130
CC FT STRAND     132    133
CC FT HELIX      136    138
CC FT TURN       141    142
CC FT HELIX      144    146
CC FT STRAND     148    152

Query Match          33.1%   Score 60; DB 1; Length 155;
Best Local Similarity 40.0%; Pred. No. 0.21;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY      3 WCTWDAPFLVTHPOGVIEGRHLDV 27
        ! : I : L : T : H | : : : ||| | 
Db       33 YCKNGGFFLRINPDGRVDGVRKSD 57

RESULT_6
FGF2_SHEEP
ID FGF2_SHEEP STANDARD PRI: 155 AA.
AC P20003;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor). (BGF) (Prostatoplin).
GN FGF2 OR FGF-2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euarcharia; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_Taxid=9940;
RN RN [1]
RP SEQUENCE FROM N.A.
RA Sutton R., Ward W.G., Raphael K.A., Cam G.R.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DDBJ databases.
[2]
RP SEQUENCE OF 9-155.
RX MEDLINE=88055577; PubMed=3678486;
RA Simpson R.J., Moritz A.L., Lloyd C.J., Fabri L.J., Nice E.C.,
RA Rubira M.R., Burgess A.W.;
RT "Primary structure of ovine pituitary basic fibroblast growth
factor.";
RL FEBS Lett. 224:128-132(1987).
CC -I FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
CC CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC -I SUBUNIT: MONOMER.
CC -I MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES
CC ARGF.
CC -----
CC -I SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration between
CC the Swiss Institute of Bioinformatics and the EMBL Outstation at its
CC European Bioinformatics Institute. There are no restrictions on the use by non-profit institutions as long as its content is in no way used or modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to licensel@isb-sib.ch).
CC -----
CC DR EMBL; J36136; AAA3J519.1; -.
CC DR PIR; S00185; S00185.
CC DR HSSP; P09038; BFE.
CC DR InterPro; IPR002209; HB/F_growthfact.
CC DR InterPro; IPR002348; IL_IL_HBGF.
CC Dr Pfam; PF00167; FGF_1.
CC PRINTS; PR00262; IL_HBGF.
```

```

DR ProDom: PD000831; HB/F-growthfact; 1.
DR SMART: SM00442; FGF; 1.
DR PROSITE: PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 9
FT CHAIN 10 155
FT SITE 45 48
FT SITE 87 90
FT BINDING 27 31
FT BINDING 116 119
SQ SEQUENCE 155 AA; 17280 MW; B5F2364BA610606D CRC64;

Query Match
Best Local Similarity 33.1%; Score 60; DB 1; Length 155;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTMDAFYLVHPGVIEGVRLVD 27
DB 33 YCKNGGFLLRHPDGRVDGVRKSD 57

RESULT 7
FGF2_MONDO
ID FGF2_MONDO STANDARD; PRT; 156 AA.
AC P48798;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Heparin-binding growth factor 2 precursor (HBGF-2) (Basic fibroblast
DE growth factor) (BFGF) (Prostatropin).
GN FGF2
OS Monodelphis domestica (Short-tailed grey opossum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Didelphimorphia; Didelphidae; Monodelphis.
RN NCBI_TaxID=13616;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE-EYE.
RX MEDLINE-94296558; PubMed-8024698;
RA Kuswitt D.F., Sabourin C.L.K., Sherburn T.E., Ley R.D.;
RA "Characterization of cDNA encoding simple fibroblast growth factor of
RA the marsupial Monodelphis domestica.";
RL DNA Cell Biol. 13:549-554(1994).
CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
CC CONCENTRATION OF THESE 2 GROWTH FACTORS.
CC -1- SUBUNIT: MONOMER.
CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES
CC ARGF.
CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15154; CAAT8854.1; ALT_INIT.
DR HSSP: P09038; 1BFF.
DR InterPro: IPR002209; HB/F-growthfact.
DR InterPro: IPR002348; IIL_HBGF.
DR Pfam: PF00167; FGF; 1.
DR PRINTS: PR00262; IILHBGF.
DR PRODOM: PD000831; HB/F-growthfact; 1.
DR SMART: SM00442; FGF; 1.
DR PROSITE: PS00247; HBGF_FGF; 1.
KW Growth factor; Mitogen; Angiogenesis; Heparin-binding.
FT PROPEP 1 9
FT CHAIN 10 156
FT BINDING 28 32
BY SIMILARITY.
HEPARIN-BINDING GROWTH FACTOR 2.
HEPARIN (POTENTIAL).

```

```

FT BINDING 117 120
SQ SEQUENCE 156 AA; 17303 MW; 7E655FCC49BF1209 CRC64;
HEPARIN (POTENTIAL).

Query Match
Best Local Similarity 32.6%; Score 59; DB 1; Length 156;
Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

OY 3 WCTMDAFYLVHPGVIEGVRLVD 27
DB 34 YCKNGGFLLRHPDGRVDGVRKSD 58

RESULT 8
SSUA_ECOLI
ID SSUA_ECOLI STANDARD; PRT; 319 AA.
AC P75853; O9R705;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative aliphatic sulfonates binding protein precursor.
DE SSUA OR B0936.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-99436146; PubMed-10506196;
RA Van der Ploeg J.R., Iwanicka-Nowicka R., Bykowski T., Hryniewicz M.,
RA Leisinger T.;
RA "The Escherichia coli ssuEADCB gene cluster is required for the
RA utilization of sulfur from aliphatic sulfonates and is regulated by
RA the transcriptional activator Cbl.";
RL J. Biol. Chem. 274:29358-29365(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-97426617; PubMed-9278503;
RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RA "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-97061202; PubMed-8905232;
RA Oshima T., Alida H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Ikeda T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Selt Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiiuchi T.;
RA "A 718-Kb DNA Sequence of Escherichia coli K-12 Genome Corresponding
RA to the 12.7-28.0 min Region on the Linkage Map.";
RL DNA Res. 3:137-155(1996).
CC -1- FUNCTION: PART OF A BINDING-PROTEIN-DEPENDENT TRANSPORT SYSTEM FOR
CC ALIPHATIC SULFONATES. PUTATIVE BINDING PROTEIN.
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: TO B.SUBTILIS SSUA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AJ237695; CAB40390.1;
DR EMBL: AE000195; AAC74022.1; ALT_INIT.

```

DR EMBL; D90732; BAA35691.1; ALT_INIT.
 DR EcoGene; EG13707; sauA.
 DR InterPro; IPR001638; SBP_bac_3.
 DR SMART; SM00652; PBPB; 1. Signal: Complete proteome.
 DR TransPort; Periplasmic; Signal: Complete proteome.
 FT SIGNAL 1 21 POTENTIAL.
 FT CHAIN 22 319 PUTATIVE ALIPHATIC SULFONATES BINDING
 FT PROTEIN.
 SO SEQUENCE 319 AA; 34557 MW; BACB0711F38D53D0 CRC64;

Query Match 29.6%; Score 53.5; DB 1; Length 319;
 Best Local Similarity 46.2%; Pred. No. 3.4;
 Matches 12; Conservative 1; Mismatches 10; Indels 3; Gaps 1;

Oy 3 WCTMDAFYLVHPGVIGVRLVDS 28
 Db 180 WAIWDPYTSALLQG--GVRLKDG 202

RESULT 9

FGF2_CHICK STANDARD: PRT: 158 AA.
 AC P48800.
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Heparin-binding growth factor 2 precursor (HBGF-2) (basic fibroblast
 growth factor) (BFGF).
 GN FGF2 OR FGF-2.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Archaeoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 CC Gallus.
 CC NCBI_TaxID=9031;

RM [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93246053; PubMed=7683281;
 RA Borja A.Z., Zeller R., Meljers C.;
 RT "Expression of alternatively spliced bfgf first coding exons and
 RT antisense mRNAs during chicken embryogenesis.";
 RL Dev. Biol. 157:110-118(1993).

CC -1- FUNCTION: THE HEPARIN-BINDING GROWTH FACTORS ARE ANGIOGENIC AGENTS
 CC IN VIVO AND ARE POTENT MITOGENS FOR A VARIETY OF CELL TYPES IN
 CC VITRO. THERE ARE DIFFERENCES IN THE TISSUE DISTRIBUTION AND
 CC CONCENTRATION OF THESE 2 GROWTH FACTORS.

CC -1- SUBUNIT: MONOMER.
 CC -1- MISCELLANEOUS: THIS PROTEIN BINDS HEPARIN MORE STRONGLY THAN DOES
 CC AREG.

CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M95707; AAA48617.1; -.
 DR HSSP; P09038; 1BFF.
 DR InterPro; IPR002209; HB/F-growthfact.
 DR InterPro; IPR002348; IIL_HBGF.
 DR Pfam; PF00167; FGF; 1.
 DR PRINTS; PR00262; IILHBGF.
 DR PRODOM; PD000831; HB/F-growthfact; 1.
 DR SMART; SM00442; FGF; 1.
 DR PROSITE; PS00247; HBGF_FGF; 1.
 DR Growth factor; Mitogen; Angiogenesis; Heparin-binding.
 FT PROPEP 1 12 BY SIMILARITY.
 FT CHAIN 13 158 HEPARIN-BINDING GROWTH FACTOR 2.
 FT BINDING 30 34 HEPARIN (POTENTIAL).
 FT BINDING 119 122 HEPARIN (POTENTIAL).
 SO SEQUENCE 158 AA; 17374 MW; 7869B864C17F1816 CRC64;

Query Match 29.3%; Score 53; DB 1; Length 158;
 Best Local Similarity 36.0%; Pred. No. 2;
 Matches 9; Conservative 6; Mismatches 10; Indels 0; Gaps 0;

Oy 3 WCTMDAFYLVHPGVIGVRLVDS 27
 Db 36 YCKNGFFLRINPDGRVDVREKSD 60

RESULT 10

PTP6_DROME STANDARD: PRT: 1462 AA.
 ID PTP6_DROME
 AC P16620.
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Protein-tyrosine phosphatase DPTP precursor (EC 3.1.3.48) (Protein-
 DE tyrosine-phosphate phosphohydrolase).
 GN PTP69D OR DPTP.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 CC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 CC NCBI_TaxID=7227;

RM [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90045860; PubMed=2554325;
 RA Streuli M., Krueger N.X., Tsai A.Y.M., Salto H.;
 RT "A family of receptor-linked protein tyrosine phosphatases in humans
 RT and Drosophila.";
 RL Proc. Natl. Acad. Sci. U.S.A. 86:8698-8702(1989).

CC -1- FUNCTION: IT IS POSSIBLE THAT DPTP IS A CELL ADHESION RECEPTOR
 CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein
 CC tyrosine + phosphate.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- SIMILARITY: CONTAINS 2 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M27699; AAA28842.1; -.
 DR PIR; B36182; B36182.
 DR HSSP; P18052; TYPO.
 DR FLYBase; FBgn0014007; Ptp69D.
 DR InterPro; IPR003861; FN_III.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003600; Ig_C2.
 DR InterPro; IPR000387; TYR_phosphatase.
 DR InterPro; IPR000242; Tyr_pp.
 DR Pfam; PF00041; fn3; 3.
 DR Pfam; PF00047; Ig; 2.
 DR Pfam; PF00102; Y_phosphatase; 2.
 DR PRINTS; PR00700; PTPYPPHPTASE.
 DR SMART; SM00060; FN3; 3.
 DR SMART; SM00410; IG_Like; 1.
 DR SMART; SM00408; IGC2; 1.
 DR SMART; SM00194; PTPc; 2.
 DR PROSITE; PS00383; TYR_PHOSPHATASE_1; 2.
 DR PROSITE; PS50056; TYR_PHOSPHATASE_2; 2.
 DR PROSITE; PS50055; TYR_PHOSPHATASE_PP; 2.
 DR Hydrolase; Receptor; Glycoprotein; Signal; Transmembrane;
 KW Cell adhesion; Immunoglobulin domain; Repeat.
 FT SIGNAL 1 23
 FT CHAIN 24 1462 PROTEIN-TYROSINE PHOSPHATASE DPTP.

FT DOMAIN 24 805 EXTRACELLULAR (POTENTIAL).
 FT TRASMEN 806 823 POTENTIAL.
 FT DOMAIN 824 1462 CYTOPLASMIC (POTENTIAL).
 FT DOMAIN 38 119 IG-LIKE C2-TYPE DOMAIN 1.
 FT DOMAIN 147 221 IG-LIKE C2-TYPE DOMAIN 2.
 FT DOMAIN 332 437 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 438 538 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 912 1165 PROTEIN-TYROSINE PHOSPHATASE 1.
 FT DOMAIN 1208 1459 PROTEIN-TYROSINE PHOSPHATASE 2.
 FT ACT_SITE 1097 1097 BY SIMILARITY.
 FT ACT_SITE 1391 1391 BY SIMILARITY.
 FT DISULFID 45 112 POTENTIAL.
 FT DISULFID 154 214 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 58 58 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 64 64 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 105 105 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 119 119 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 191 191 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 235 235 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 429 429 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 451 451 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 516 516 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 613 613 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 701 701 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 755 755 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1462 AA; 167411 MW; F8091D69E88230EB CRC64;

Query Match 28.7%; Score 52; DB 1; Length 1462;
 Best Local Similarity 32.0%; Pred. No. 24;
 Matches 8; Conservative 6; Mismatches 11; Indels 0; Gaps 0;

Oy 1 FGMCTWDAFLTVHPGVIEGRHL 25
 Db 1058 YHVLTKDFMAEPHGHGKIKFRQI 1082

RESULT 11
 H181_ANASP STANDARD; PRT; 353 AA.
 AC Q8YV69; 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Histidinol-phosphate aminotransferase 1 (EC 2.6.1.9) (Imidazole
 acetol-phosphate transaminase 1).
 GN H181 OR ALR2092.
 OS Anabaena sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
 OX NCBI_TaxID=103690;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21595285; PubMed=11759840;
 RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
 RA Watanabe A., Iriyuchi M., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kohara M., Matsumoto M., Matsuo A., Muraki A.,
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
 RA Yasuda M., Tabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium Anabaena sp. strain PCC 7120.";
 RL DNA Res. 8:205-213(2001).
 CC -1- CATALYTIC ACTIVITY: L-histidinol-phosphate + 2-oxoglutarate = 3-
 CC (imidazol-4-yl)-2-oxopropyl phosphate + L-glutamate.
 CC -1- Cofactor: Pyridoxal phosphate (By similarity).
 CC -1- PATHWAY: Histidine biosynthesis; seventh step.

CC -1- SIMILARITY: BELONGS TO CLASS-II OF PYRIDOXAL-PHOSPHATE-DEPENDENT
 CC AMINOTRANSFERASES. HISTIDINOL-PHOSPHATE AMINOTRANSFERASES
 CC SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; AP003588; BAB73791.1; .
 DR InterPro; IPR004839; AminoTransf1/2.
 DR InterPro; IPR001917; NHTransf_2.
 DR Pfam; PF00155; amntran.1.2; 1.
 DR TIGRfam; TIGR01141; hisc.1.
 DR PROSITE; PS00599; AA-TRANSFER_CLASS_2; 1.
 KW Histidine biosynthesis; Transferrase; Aminotransferase;
 KW Pyridoxal phosphate; Complete proteome.
 FT BINDING 211 211 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
 SQ SEQUENCE 353 AA; 39105 MW; 3891CF87E990A4E CRC64;

Query Match 28.5%; Score 51.5; DB 1; Length 353;
 Best Local Similarity 50.0%; Pred. No. 7.1;
 Matches 11; Conservative 2; Mismatches 6; Indels 3; Gaps 1;

Oy 2 GWCMTWDA---FYLTVHPGVIE 20
 Db 282 GFCVWDSQANFLTTPPGQNAE 303

RESULT 12
 E312_ADE03 STANDARD; PRT; 106 AA.
 AC P11319; 01-JUL-1989 (Rel. 11, Created)
 DT 01-JUL-1989 (Rel. 11, Last sequence update)
 DE 01-AUG-1992 (Rel. 23, Last annotation update)
 DE Probable early E3 12.1 kDa glycoprotein.

OS Human adenovirus type 3.
 OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
 OX NCBI_TaxID=43659;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87219876; PubMed=3582978;
 RA Signes C., Akusjaervi G., Pettersson U.;
 RT "Region E3 of human adenoviruses; differences between the oncogenic
 RT adenovirus-3 and the non-oncogenic adenovirus-2.";
 RL Gene 50:173-184(1986).
 CC -1- FUNCTION: NOT YET KNOWN.
 CC -1- SIMILARITY: TO EQUIVALENT PROTEIN IN OTHER ADENOVIRUSES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; M15952; AAA42481.1; .
 DR PIR; B29500; ERAD31.
 KW Early protein; Glycoprotein.
 FT CARBOHYD 77 77 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 106 AA; 12124 MW; C0374ACAB6CECEC0 CRC64;

Query Match 28.2%; Score 51; DB 1; Length 106;
 Best Local Similarity 56.2%; Pred. No. 2.6;
 Matches 9; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Oy 14 HPGGVIEGVRLHVDG 29
 Db 1111 111 111

Db 40 HPOGPAHGVRITIEGG 55

RESULT 13

PIP-NEIGO STANDARD; PRT; 310 AA.

AC P42786; 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Proline iminopeptidase (EC 3.4.11.5) (PIP) (Prolyl aminopeptidase) (PAP).

GN PIP.

OS Neisseria gonorrhoeae.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=485;

RM [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.

RC STRAIN-MS11 / MS01-1X;

RX MEDLINE=95020651; PubMed=7934933;

RA Albertson N.H., Koomey M.;

RT "Molecular cloning and characterization of a proline iminopeptidase gene from Neisseria gonorrhoeae."

RL Mol. Microbiol. 9:1203-1211(1993).

CC -1- FUNCTION: HYDROLYSES PEPTIDES HAVING THE STRUCTURE PRO-Y-Z TO YIELD FREE PROLINE. ALSO HYDROLYSES THE DIPEPTIDE PRO-GLY.

CC -1- CATALYTIC ACTIVITY: Release of a N-terminal proline from a peptide.

CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S33.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z25461; CAAB0948.1; -.

DR HSSP: O32449; 10TR.

DR MEROPS: S33.001; -.

DR InterPro: IPR000073; Abhydrolase.

DR InterPro: IPR002410; Pro_aminopeptase.

DR InterPro: IPR000379; Ser_estr_silte.

DR Pfam: PF00561; abhydrolase_1.

DR PRINTS: PR00793; PROAMINOPTASE.

DR TIGRfams: TIGR01249; pro_aminopep_1; 1.

KW Hydrolase; Amino-peptidase.

FT ACT_SITE 107 107 NUCLEOPHILE (BY SIMILARITY).

FT ACT_SITE 260 260 BY SIMILARITY.

FT ACT_SITE 287 287 PROTON DONOR (BY SIMILARITY).

SO SEQUENCE 310 AA; 34791 MW; DOE7ABBA80BAAE0 CRC64;

Query Match 27.3%; Score 49.5; DB 1; Length 310;

Best Local Similarity 23.7%; Pred. No. 12;

Matches 9; Conservative 7; Mismatches 11; Indels 11; Gaps 1;

Oy 3 WCTNDAFYITVHPOGVIEGV-----RHLDGCG 29

Db 195 WADWESYLRFEPEGVDEDAVSAIAIRLENHYFVNGG 232

RESULT 14

YAEI-ECOLI STANDARD; PRT; 191 AA.

AC P31546; 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Hypothetical protein yaeI.

GN YAEI OR B0200 OR Z0212 OR ECS0202.

OS Escherichia coli, and

OS Escherichia coli O157:H7.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

CC Escherichia

OX NCBI_TaxID=562, 83334;

RM [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12;

RA Miyamoto K.;

RL Submitted (APR-1993) to the EMBL/Genbank/DBJ databases.

RM [2]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12."

RL Science 277:1453-1474(1997).

RM [3]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / W3110;

RA Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T., Yamamoto Y., Inokuchi H., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S., Mizuno T., Makino K., Nakata A., Yura T., Sampei G., Mizobuchi K.;

RT "Systematic sequencing of the Escherichia coli genome: analysis of the 4.0 - 6.0 min (189,987 - 281,415bp) region."

RL Submitted (FEB-1996) to the EMBL/Genbank/DBJ databases.

RM [4]

RP SEQUENCE FROM N.A.

RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E., Davis K., Federspiel N., Hyman R., Kaiman S., Komp C., Kurd O., Laeskari D., Lew H., Lin D., Namath A., Oefner P., Roberts D., Davis R.W.;

RL Submitted (SEP-1996) to the EMBL/Genbank/DBJ databases.

RM [5]

RP SEQUENCE FROM N.A.

RC STRAIN-O157:H7 / RIMD 0509952;

RA Ohnishi M., Murata T., Nakayama K., Kuhara S., Hattori M., Kurokawa K., Yasunaga T., Makino K., Shinagawa H., Hayashi T.;

RT "Development of primer sets for direct sequence determination of all the ribosomal operons of Escherichia coli."

RL Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.

RM [6]

RP SEQUENCE FROM N.A.

RC STRAIN-O157:H7 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551.

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D., Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A., Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L., Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K., Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C., Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."

RL Nature 409:529-533(2001).

RM [7]

RP SEQUENCE FROM N.A.

RC STRAIN-O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K., Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T., Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T., Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12."

RL DNA Res. 8:11-22(2001).

CC -1- SIMILARITY: STRONG, TO H. INFLUENZAE HI0621.1.

CC -1- SIMILARITY: TO M. JANNASCHII MJO015.

CC -1- SIMILARITY: TO THE N-TERMINAL PART OF E. COLI HISB, THE HISTIDINOL PHOSPHATE PHOSPHATASE DOMAIN.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D15061; BAA03661.1; ALT_INIT.
 DR EMBL: AE000129; AAC7331.1; -
 DR EMBL: D83536; BAA77877.1; -
 DR EMBL: U70214; AAB08628.1; -
 DR EMBL: AB035926; BAA93568.1; -
 DR EMBL: AE005195; AAG54502.1; -
 DR EMBL: AP002550; BAB33625.1; -
 DR EMBL: EG11736; yaed.
 DR EcoGene: IPR001454; Hlg_nase/hydrlase.
 DR InterPro: IPR004446; Hstcl_phs_rel.
 DR Pfam: PF00702; Hydrolase: 1.
 DR TIGRfams: TIGR00213; hstcl_phs_rel; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 191 AA; 21294 MW; E7814B34A23128FA CRC64;
 SQ

Query Match 27.1%; Score 49; DB 1; Length 191;
 Best Local Similarity 28.6%; Pred. No. 8.8;
 Matches 10; Conservative 3; Mismatches 12; Indels 10; Gaps 1;

OY 3 WCTWD-----AFYLVHPGVIEGVRHLVD 27
 DB 74 WMDWSLADRDVDLDGIYCPHFGSVSEEFQVCD 108

RESULF 15
 LIPB_SYNY3
 ID LIPB_SYNY3 STANDARD: PRT; 227 AA.
 AC P74519;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lipote-protein ligase B (EC 6.-.-) (Lipote biosynthesis protein
 B).
 GN LIPB OR SLR0994.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
 RA Hosonouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
 RA Okumura S., Shimo S., Takeuchi C., Wada T., Watanabe A.,
 RA Yamada M., Yasuda M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
 RT entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).
 CC -!- FUNCTION: INVOLVED IN THE ATTACHMENT OF LIPOYL GROUPS TO PROTEINS,
 CC BY CREATING AN AMIDE LINKAGE THAT JOINS THE FREE CARBOXYL GROUP OF
 CC LIPOIC ACID TO THE EPSILON-AMINO GROUP OF A SPECIFIC LYSINE
 CC RESIDUE IN LIPOYLATED PROTEINS (BY SIMILARITY).
 CC -!- PATHWAY: Lipote biosynthesis.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE LIPB FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D90915; BAA18623.1; -
 DR InterPro: IPR004143; BPL_LipA_LipB.

DR InterPro: IPR000544; Lipote_B.
 DR Pfam: PF03099; BPL_LipA_LipB; 1.
 DR ProDom: PD006086; Lipote_B; 1.
 DR TIGRfams: TIGR00214; LipB; 1.
 DR PROSITE: PS01313; LIPB; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 227 AA; 25640 MW; 0D89E0C781030A46 CRC64;

Query Match 27.1%; Score 49; DB 1; Length 227;
 Best Local Similarity 39.3%; Pred. No. 10;
 Matches 11; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

OY 2 GWCWDAPYLVHPGVIEGVRHLVDG 29
 DB 150 GWVTCGFALNICPD--LEGFSHIVPCG 175

Search completed: March 27, 2003, 10:04:15
 Job time : 5.93902 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: March 27, 2003, 10:01:36 ; Search time 16.8293 Seconds
(without alignments)
367.301 Million cell updates/sec

Title: US-09-675-208-1
Perfect score: 181
Sequence: 1 FGMCTMDAFYLVHPQGVIEGVRHLVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protist:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	181	100.0	784	10 Q92T62	Q92T62 cucumis sat
2	153	84.5	816	10 Q942T6	Q942T6 oryza sativ
3	147	81.2	798	10 Q8VNM6	Q8VNM6 pisum sativ
4	144	79.6	783	10 Q8FND9	Q8FND9 arabidopsis
5	124	68.5	863	10 Q8RW08	Q8RW08 stachys aff
6	117	64.6	857	10 Q9SXX0	Q9SXX0 phaseolus a
7	111	61.3	765	10 Q43408	Q43408 brassica ol
8	111	61.3	853	10 Q93XK2	Q93XK2 pisum sativ
9	108	59.7	757	10 Q39466	Q39466 cicer ariet
10	108	59.7	773	10 Q9SCM1	Q9SCM1 hordeum vul
11	108	59.7	773	10 Q9SCM1	Q9SCM1 arabidopsis
12	108	59.7	773	10 Q94A08	Q94A08 arabidopsis
13	106	58.6	807	10 Q9SYA4	Q9SYA4 arabidopsis
14	101	55.8	779	10 Q9M4M7	Q9M4M7 persea amer
15	99	54.7	844	10 Q8RX87	Q8RX87 arabidopsis
16	82	45.3	1170	10 Q9LFZ7	Q9LFZ7 arabidopsis

17	72	39.8	371	10 Q04607	Q04607 arabidopsis
18	64.5	35.6	374	16 Q9KRP2	Q9KRP2 vibrio chol
19	63	34.8	702	2 Q93CM6	Q93CM6 bifidobacte
20	61	33.7	198	16 Q92KR4	Q92KR4 rhizobium m
21	60.5	33.4	674	17 Q96XG2	Q96XG2 sulfolobus
22	60	33.1	114	4 Q16443	Q16443 homo sapien
23	60	33.1	114	4 Q00527	Q00527 homo sapien
24	60	33.1	130	6 Q77767	Q77767 canis fam11
25	60	33.1	153	11 Q925A3	Q925A3 mus musculu
26	60	33.1	196	4 P78443	P78443 homo sapien
27	60	33.1	451	5 Q21801	Q21801 caenorhabdi
28	57.5	31.8	649	17 Q97094	Q97094 sulfolobus
29	57	31.5	170	11 Q60487	Q60487 cavia porce
30	56.5	31.2	221	5 Q9GPI2	Q9GPI2 crassostrea
31	56.5	31.2	373	16 Q82B05	Q82B05 versinia pe
32	54	29.8	106	12 Q8UW75	Q8UW75 stimen aden
33	54	29.8	390	10 Q43811	Q43811 pelargonium
34	54	29.8	390	10 Q43812	Q43812 pelargonium
35	54	29.8	489	16 P93772	P93772 pelargonium
36	53.5	29.6	319	16 Q8XDP8	Q8XDP8 escherichia
37	53	29.3	557	16 Q9KEU2	Q9KEU2 bacillus ha
38	52.5	29.0	265	2 P72162	P72162 pseudomonas
39	52.5	29.0	265	16 Q9HTB8	Q9HTB8 pseudomonas
40	52	28.7	237	10 Q9C7N3	Q9C7N3 arabidopsis
41	52	28.7	619	16 Q9K228	Q9K228 chlamydia p
42	52	28.7	619	16 Q92935	Q92935 chlamydia p
43	52	28.7	641	17 Q978X5	Q978X5 thermoplas
44	52	28.7	650	4 Q96ED3	Q96ED3 homo sapien
45	52	28.7	714	4 Q8WU76	Q8WU76 homo sapien

ALIGNMENTS

RESULT 1	Q92T62	PRELIMINARY:	PRT:	784 AA.
ID	Q92T62			
AC	Q92T62			
DT	01-MAY-1999 (TREMBLrel. 10, Created)			
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)			
DT	01-MAR-2002 (TREMBLrel. 20, Last annotation update)			
DE	Raffinose synthase (EC 2.4.1.82).			
GN	RFS.			
OS	Cucumis sativus (Cucumber).			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;			
OC	eurostids I; Cucurbitales; Cucurbitaceae; Cucumis.			
OX	NCBI_TaxID=3659;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAF;			
RA	Onsumi C., Nozaki J., Kida T.;			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF073744; AAD02832.1; -			
KW	Glycosyltransferase; Transferase.			
SO	SEQUENCE 784 AA; 86920 MW; 3806A91F0908933 CRC64;			
Query Match	100.0%;	Score 181;	DB 10;	Length 784;
Best Local Similarity	100.0%;	Pred. No. 7e-17;		
Matches	30;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;
QY	1 FGMCTMDAFYLVHPQGVIEGVRHLVDGCG 30			
DB	215 FGMCTMDAFYLVHPQGVIEGVRHLVDGCG 244			
RESULT 2	Q942T6	PRELIMINARY:	PRT:	816 AA.
ID	Q942T6			
AC	Q942T6			
DT	01-DEC-2001 (TREMBLrel. 19, Created)			
DT	01-DEC-2001 (TREMBLrel. 19, Last sequence update)			
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)			

Query	Best Local Similarity	Score	DB	Length	Indels	Gaps
DE Putative raffinose synthase.	84.5%	153	DB 10	816	0	0
GN P0583608.2.	83.3%	13	DB 10	816	0	0
OS Oryza sativa (Rice).	81.2%	147	DB 10	798	0	0
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;	80.0%	12	DB 10	798	0	0
OC Spematophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;	79.8	11	DB 10	798	0	0
OC Euhartioideae; Oryzaceae; Oryza.	78.3	10	DB 10	798	0	0
NCBI_TaxID=4530;	77.8	9	DB 10	798	0	0
RN [1]	77.3	8	DB 10	798	0	0
RP SEQUENCE FROM N.A.	76.8	7	DB 10	798	0	0
RC STRAIN=CV. NIPPONBARE;	76.3	6	DB 10	798	0	0
RA Sasaki T., Matsunoto T., Yamamoto K.;	75.8	5	DB 10	798	0	0
RT "Oryza sativa nipponbare(Ga3) genomic DNA, chromosome 1, PAC	75.3	4	DB 10	798	0	0
RT clone:00583608."	74.8	3	DB 10	798	0	0
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.	74.3	2	DB 10	798	0	0
DR EMBL; AP003282; BAB64768.1; -	73.8	1	DB 10	798	0	0
SQ SEQUENCE 816 AA; 89588 MW; B316DEF3566C5178 CRC64;	73.3	0	DB 10	798	0	0
Query Match	84.5%	153	DB 10	816	0	0
Best Local Similarity	83.3%	13	DB 10	816	0	0
Matches 25; Conservative 1; Mismatches 4; Indels 0; Gaps 0						
DB 227 FGMCTWDAFYLVHPHOGVIEGVRHLVDGCG 256						
1 FGMCTWDAFYLVHPHOGVIEGVRHLVDGCG 30						
08VWN6 PRELIMINARY; PRT; 798 AA.						
08VWN6						
AC 08VWN6						
DT 01-MAR-2002 (TREMblrel. 20, Created)						
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)						
DT 01-MAR-2002 (TREMblrel. 20, Last annotation update)						
DE Raffinose synthase (EC 2.4.1.82).						
GN RFS.						
OS Pisum sativum (Garden pea).						
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;						
OC eustosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.						
OK NCBI_TaxID=3888;						
RN [1]						
RP SEQUENCE FROM N.A.						
RC STRAIN=CV. WUNDER VON KELYDON; TISSUE=SEED;						
RA Peterbauer T., Mach L., Mucha J., Richter A.;						
RT "Molecular characterization of raffinose synthase from pea (Pisum						
RT sativum L.) seeds."						
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.						
DR EMBL; AJ426475; CAD20127.2; -						
KW Transferrase; Glycosyltransferase.						
SQ SEQUENCE 798 AA; 88717 MW; 803F3BD5BF8617B7 CRC64;						
Query Match	81.2%	Score 147;	DB 10;	Length 798;		
Best Local Similarity	80.0%	Pred. No. 12;				
Matches 24; Conservative 2; Mismatches 4; Indels 0; Gaps 0						
DB 235 FGMCTWDAFYLVHPHOGVIEGVRHLVDGCG 30						
1 FGMCTWDAFYLVHPHOGVIEGVRHLVDGCG 30						
09FND9 PRELIMINARY; PRT; 783 AA.						
09FND9						
AC 09FND9						
DT 01-MAR-2001 (TREMblrel. 16, Created)						
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)						
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)						
DE Raffinose synthase protein.						
GN MP012.13.						
OS Arabidopsis thaliana (Mouse-ear cress).						
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;						
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;						
OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.						

```

OX NCBI_TaxID=3702:
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA:
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned pl clones.";
RL DNA Res. 4:291-300(1997).
RN [2]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamita A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Natsuka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamita A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Natsuka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006702; BAH1595.1; -
DR EMBL; AY062781; AAL32859.1; -
DR EMBL; AY061645; AAM10207.1; -
SQ SEQUENCE 783 AA; 86237 MW; 3C37D1D7871888AC CRC64;

```

Query Match	79.6%	Score 144	DB 10	Length 783
Best Local Similarity	80.0%	Pred. No. 1.le-11		
Matches	24	Conservative	3	Mismatches 3
				Indels 0
				Gaps 0
QY	1	FGWCTWDAFYLTVHPQGVIEGVRHLVDGCG	30	
Db	224	FGWCTWDAFYLTVHPDGVHKGKCLVDGCG	253	
RESULT 5				
Q8RM08				
Q8RM08	PRELIMINARY	PRT	863 AA	
AC	Q8RM08			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Stachyose synthase (EC 2.4.1.67).			
GN	STS.			
OS	stachys affinis.			
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;			
OC	Asteridae; euasterids I; Lamiales; Lamiaceae; Stachys.			
OX	NCBI_TaxID=168825;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=LEAF;			
RA	Pesch M., Schmitz K.;			
RT	"Molecular cloning of a cDNA encoding for stachyose synthase from			
RT	Stachys sieboldii."			
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL, AJ344091, CAC86963.1; -			
KW	Transferase; Glycosyltransferase.			
SQ	SEQUENCE 863 AA; 95227 MW; DD96FE66099DDEPD CRC64;			
Best Match	68.5%	Score 124	DB 10	Length 863;
Best Local Similarity	69.0%	Pred. No. 7.8e-09;		
Matches	20	Conservative	2	Mismatches 7
				Indels 0
				Gaps 0
QY	1	FGWCTWDAFYLTVHPQGVIEGVRHLVDGCG	29	

Db 224 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 252

RESULT 6

Q9SBZ0 PRELIMINARY: PRT: 857 AA.
 ID 09SBZ0:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Galactinol-4-epimerase galactosyltransferase (EC 2.4.1.67).
 OS Phaseolus angularis (Adzuki bean) (Vigna angularis).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Vigna.
 NCBI_TaxID=3914;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20117502; PubMed=10652123;
 RA Peterbauer T., Mucha J., Mayer U., Popp M., Gloessel J., Richter A.;
 RT "Stachyose synthesis in seeds of adzuki bean (Vigna angularis).";
 RT molecular cloning and functional expression of stachyose synthase.";
 RL Plant J. 20:509-518(1999).
 DR EMBL: Y19024; CAB64563.1; -;
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 857 AA; 94949 MW; 85248C4B81165679 CRC64;

Query Match 64.6%; Score 117; DB 10; Length 857;
 Best Local Similarity 65.5%; Pred. No. 7.4e-08;
 Matches 19; Conservative 4; Mismatches 6; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 29
 Db 221 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 249

RESULT 7

Q43408 PRELIMINARY: PRT: 765 AA.
 ID 043408:
 AC 043408:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Putative imbibition protein.
 OS Brassica oleracea (Cauliflower).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Brassicales; Brassicaceae; Brassica.
 NCBI_TaxID=3712;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. ALPHA BALOMA; TISSUE=ROOT TIPS;
 RA Fujikura Y., Karssen C.K.;
 RT "Cauliflower cDNA encoding a putative imbibition protein.";
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X79330; CAA55893.1; -;
 SQ SEQUENCE 765 AA; 84084 MW; 73398603048E9B58 CRC64;

Query Match 61.3%; Score 111; DB 10; Length 765;
 Best Local Similarity 65.5%; Pred. No. 4.5e-07;
 Matches 19; Conservative 3; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 29
 Db 198 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 226

RESULT 8

Q93XK2 PRELIMINARY: PRT: 853 AA.
 ID 093XK2:
 AC 093XK2:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)

DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
 DE Stachyose synthase (EC 2.4.1.67).
 GN STS1.

OS Pisum sativum (Garden pea).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciae; Pisum.
 NCBI_TaxID=3888;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. WUNDER VON KLEVEDON; TISSUE=SEED;
 RA Peterbauer T.;
 RT "Molecular cloning of a cDNA encoding for stachyose synthase from pea
 seeds.";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ311087; CAC38094.1; -;
 KW Glycosyltransferase; Transferase.
 SQ SEQUENCE 853 AA; 95890 MW; CB10F18C8D37B38C CRC64;

Query Match 61.3%; Score 111; DB 10; Length 853;
 Best Local Similarity 62.1%; Pred. No. 5.1e-07;
 Matches 18; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 29
 Db 221 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 249

RESULT 9

Q39466 PRELIMINARY: PRT: 357 AA.
 ID 039466:
 AC 039466:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Seed imbibition protein (Fragment).
 GN SIP.
 OS Cicer arietinum (Chickpea) (Garbanzo).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Ciceraceae; Cicer.
 NCBI_TaxID=3827;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cervantes E.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: X95875; CAA65125.1; -;
 FT NON_TER 1 1
 FT NON_TER 357 357
 SQ SEQUENCE 357 AA; 39465 MW; E8CD30399E1B33EE CRC64;

Query Match 59.7%; Score 108; DB 10; Length 357;
 Best Local Similarity 62.1%; Pred. No. 5.2e-07;
 Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 29
 Db 31 FGMCTWDAFYLTVPAGIIMGVKEPSSDGG 59

RESULT 10

Q40077 PRELIMINARY: PRT: 757 AA.
 ID 040077:
 AC 040077:
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Seed imbibition protein.
 GN S1P1.
 OS Hordeum vulgare (Barley).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;

OC Triliceae; Hordeum.
 OX NCBI_TaxID=4513;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-HIMALAYA:
 RA Heck G.R., Dorsett C., Ho T.-H.D.:
 RT "Cloning and characterization of a gene, Slp1, associated with seed
 RL "Imbibition in barley.";
 DR EMBL: M77475; AAA32975.1; -;
 SQ SEQUENCE 757 AA; 82133 MW; EA9E7B771AABCCF CRC64;

Query Match 59.7%; Score 108; DB 10; Length 757;
 Best Local Similarity 62.1%; Pred. No. 1.2e-06;
 Matches 18; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 205 FGMCTWDATYLVHPQVIEGVRLHVDG 233

RESULT 11

O9SCM1 PRELIMINARY; PRT; 773 AA.

AC O9SCM1:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
 DE Imbibition protein homolog.
 GN T8H10.120.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Benes V., Reichmann S., Borkova D., Ansgorge W., Mewes H.W., Lemcke K.,
 RA Mayer K.F.X., Quetler F., Salanoubat M.;
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AL13246; CAB6109.1; -;
 SQ SEQUENCE 773 AA; 85143 MW; 0852F9E67952C8D3 CRC64;

Query Match 59.7%; Score 108; DB 10; Length 773;
 Best Local Similarity 62.1%; Pred. No. 1.2e-06;
 Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 200 FGMCTWDATYLVHPQVIEGVRLHVDG 228

RESULT 12

O94A08 PRELIMINARY; PRT; 773 AA.

AC O94A08:
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DE 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Putative imbibition protein.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Pham P.K., Banh J., Banno F., Dale J.M.,
 RA Goldsmith A.D., Jiang P.X., Lee J.M., Onodera C.S., Quach H.L.,

RA Tang C., Toriumi M., Yamamura Y., Yu G., Yu S., Bowser L.,
 RA Caminci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J., Jones T.,
 RA Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Koesema E., Lam B.,
 RA Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M., Palm C.J.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Southwick A., Tracy S.E.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene T8H10.120/AT3g57520 (GI:6706423).";
 RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AY050772; AAK92707.1; -;
 SQ SEQUENCE 773 AA; 85171 MW; 18BD9E67952C801 CRC64;

Query Match 59.7%; Score 108; DB 10; Length 773;
 Best Local Similarity 62.1%; Pred. No. 1.2e-06;
 Matches 18; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 200 FGMCTWDATYLVHPQVIEGVRLHVDG 228

RESULT 13

O9SYJ4 PRELIMINARY; PRT; 807 AA.

AC O9SYJ4:
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DE 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE Putative raffinose synthase OR seed IMBIBITION protein.
 GN T7B11.23 OR AT4G01970.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. COLUMBIA:
 RA Huang E.N., Nascimento L., de la Bastide M., Habermann K., Vil M.D.,
 RA Preston R.R., Spiegel L.A., See L.H., Shah R., Matero A.,
 RA O'Shaughnessy A., Rodriguez M., Shekher M., Swaby I., Schutz K.,
 RA Parnell L.D., Dedhia N.N., McCombie W.R.;
 RT "Arabidopsis thaliana BAC T7B11 from chromosome IV near 10 cM.";
 RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Spiegel L.A., Huang E.N., Nascimento L.U., de la Bastide M., Vil M.D.,
 RA Preston R.R., Matero A., Shah R., O'Shaughnessy A., Rodriguez M.,
 RA Shekher M., Schutz K., See L.H., Swaby I., Habermann K., Dedhia N.N.,
 RA Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC007138; AAD22659.1; -;
 DR EMBL: AL161493; CAB80690.1; -;
 SQ SEQUENCE 807 AA; 90122 MW; 963DCD5A827B338B CRC64;

Query Match 58.6%; Score 106; DB 10; Length 807;
 Best Local Similarity 62.1%; Pred. No. 2.4e-06;
 Matches 18; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

OY 1 FGMCTWDATYLVHPQVIEGVRLHVDG 29
 DB 235 FGMCTWDATYLVHPQVIEGVRLHVDG 263

RESULT 14

O9M4M7 PRELIMINARY; PRT; 779 AA.

AC O9M4M7:
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Putative seed imbibition protein.
 GN SIP.
 OS Persa americana (Avocado).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
 OX NCBI_TaxID=3435;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV, HASS; TISSUE=MESOCARP;
 RA Zamorano J.P., Evans A.D., Dopico B., Lowe A.L., Wilson I.D.,
 RT Merodio C., Grierison D.;
 RT "Isolation and characterization of cDNAs for mRNAs regulated during
 cold storage of avocado (Persea americana Mill.) fruit."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ133148; CAB77245.1;
 SQ SEQUENCE 779 AA: 85368 MW: C3A8B43160316785 CRC64;

Query Match 55.8%; Score 101; DB 10; Length 779;
 Best Local Similarity 58.6%; Pred. No. 1.2e-05;
 Matches 17; Conservative 4; Mismatches 8; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPQGYEGVRLVDGG 29
 |||||
 DB 203 FGMCTWDAFYTDVDEGVEGGLKSLSGG 231

RESULT 15

O8RX87 PRELIMINARY; PRT; 844 AA.
 AC O8RX87.
 DT 01-JUN-2002 (TREMBLrel. 21, Created)
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE AT5G20250/ES024.140.
 OS Arabidopsis thaliana (Mouse-ear cress).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Shun P., Chen H., Cheuk R., Kim C.J., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Miranda A., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Tortum M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY090237; AAL0901.1;
 SQ SEQUENCE 844 AA: 94215 MW: 2AC8AB0EA43F8056 CRC64;

Query Match 54.7%; Score 99; DB 10; Length 844;
 Best Local Similarity 58.6%; Pred. No. 2.4e-05;
 Matches 17; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

OY 1 FGMCTWDAFYLTVHPQGYEGVRLVDGG 29
 |||||
 DB 301 FGMCTWDAFYDEVTQGVAGLKLKSLAAG 329

Search completed: March 27, 2003, 10:05:56
 Job time : 20.8293 secs

GenCore version 5.1.4-p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 27, 2003, 10:04:25 ; Search time 6.95122 seconds
(without alignments)
253.439 Million cell updates/sec

Title: US-09-675-208-1
Perfect score: 181
Sequence: 1 FGMCTWDAFLYTHPQGVIEGVRLVDGCG 30

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 237916 seqs, 58723674 residues

Total number of hits satisfying chosen parameters: 237916

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCR_NEW_PUB pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep:*
6: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB pep:*
7: /cgn2_6/ptodata/2/pubpaa/PCFUS_PUBCOMB pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB pep:*
11: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep:*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep:*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep:*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	33.1	134	9 US-09-901-938-24	Sequence 24, Appl
2	60	33.1	145	9 US-09-940-601-3	Sequence 3, Appl
3	60	33.1	146	9 US-10-131-965-3	Sequence 3, Appl
4	60	33.1	146	9 US-10-131-965-5	Sequence 5, Appl
5	60	33.1	146	9 US-10-108-195-2	Sequence 2, Appl
6	60	33.1	146	9 US-10-108-195-3	Sequence 3, Appl
7	60	33.1	146	9 US-10-108-195-4	Sequence 4, Appl
8	60	33.1	146	9 US-10-108-195-7	Sequence 7, Appl
9	60	33.1	146	10 US-09-802-365-2	Sequence 2, Appl
10	60	33.1	146	10 US-09-802-365-4	Sequence 2, Appl
11	60	33.1	146	10 US-09-771-302-2	Sequence 2, Appl
12	60	33.1	146	10 US-09-886-856-2	Sequence 2, Appl
13	60	33.1	146	10 US-09-886-856-4	Sequence 4, Appl
14	60	33.1	147	9 US-09-820-596-8	Sequence 8, Appl
15	60	33.1	150	12 US-10-016-447-8	Sequence 8, Appl
16	60	33.1	155	9 US-10-081-347-28	Sequence 28, Appl
17	60	33.1	155	9 US-10-108-195-1	Sequence 1, Appl
18	60	33.1	155	9 US-09-902-460-2	Sequence 2, Appl
19	60	33.1	155	10 US-09-822-485-5	Sequence 5, Appl

20	60	33.1	155	10 US-09-802-365-6	Sequence 6, Appl
21	60	33.1	155	10 US-09-802-365-8	Sequence 8, Appl
22	60	33.1	155	10 US-09-251-263-10	Sequence 10, Appl
23	60	33.1	155	10 US-09-425-021-10	Sequence 10, Appl
24	60	33.1	155	10 US-09-886-856-6	Sequence 6, Appl
25	60	33.1	155	10 US-09-886-856-8	Sequence 8, Appl
26	60	33.1	155	10 US-09-749-7288-7	Sequence 7, Appl
27	60	33.1	158	10 US-09-826-210-2	Sequence 2, Appl
28	60	33.1	159	10 US-09-934-706-2	Sequence 2, Appl
29	60	33.1	210	10 US-09-902-773A-4	Sequence 4, Appl
30	60	33.1	333	9 US-10-108-195-6	Sequence 6, Appl
31	60	33.1	501	10 US-09-934-706-4	Sequence 4, Appl
32	53.5	29.6	333	10 US-09-815-242-10116	Sequence 10116, A
33	52	28.7	313	10 US-09-788-626-19	Sequence 19, Appl
34	51.5	28.5	280	10 US-09-815-242-11714	Sequence 11714, A
35	49	27.1	140	9 US-10-131-965-2	Sequence 2, Appl
36	48.5	26.8	161	10 US-09-893-737-288	Sequence 288, App
37	48.5	26.8	497	9 US-09-896-896A-1	Sequence 1, Appl
38	48.5	26.8	497	10 US-09-753-126-1	Sequence 1, Appl
39	48.5	26.8	592	10 US-09-753-126-4	Sequence 4, Appl
40	48	26.5	141	9 US-09-901-938-18	Sequence 18, Appl
41	48	26.5	207	10 US-09-822-485-19	Sequence 19, Appl
42	48	26.5	207	10 US-09-750-963-6	Sequence 6, Appl
43	48	26.5	207	10 US-09-805-805-5	Sequence 5, Appl
44	48	26.5	207	10 US-09-805-805-7	Sequence 7, Appl
45	48	26.5	207	10 US-09-817-814-24	Sequence 24, Appl

ALIGNMENTS

RESULT 1
US-09-901-938-24
Sequence 24, Application US/09901938
Patent No. US20020156001A1
GENERAL INFORMATION:
APPLICANT: ECONS, Michael
APPLICANT: WHITE, Kenneth
APPLICANT: STROM, Tim
APPLICANT: MEITNER, Thomas
TITLE OF INVENTION: NOVEL FIBROBLAST GROWTH FACTOR (FGF23) AND METHODS FOR USE
FILE REFERENCE: 035884-5001
CURRENT APPLICATION NUMBER: US/09/901,938
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/219,137
PRIOR FILING DATE: 2000-07-19
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 24
LENGTH: 134
TYPE: PRT
ORGANISM: Homo Sapiens
US-09-901-938-24

Query Match 33.1% Score 60; DB 9; Length 134;
Best Local Similarity 40.0%; Pred. No. 0.5;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

CY 3 WCTMDAFLYTHPQGVIEGVRLVD 27
DB 12 YCKNGCFELRIHPDGVRGVRKSD 36

RESULT 2
US-09-940-601-3
Sequence 3, Application US/09940601
Publication No. US20030004319A1
GENERAL INFORMATION:
APPLICANT: Basilio, Claudio
APPLICANT: Dellio, Pasquale
TITLE OF INVENTION: MAMMALIAN GROWTH FACTOR
FILE REFERENCE: 5986/13586-USC
CURRENT APPLICATION NUMBER: US/09/940,601

```
; CURRENT FILING DATE: 2001-08-27
; PRIOR APPLICATION NUMBER: 08/775,567
; PRIOR FILING DATE: 1996-12-31
; PRIOR APPLICATION NUMBER: 08/056,482
; PRIOR FILING DATE: 1993-05-03
; PRIOR APPLICATION NUMBER: 07/806,771
; PRIOR FILING DATE: 1991-12-06
; PRIOR APPLICATION NUMBER: 07/177,506
; PRIOR FILING DATE: 1988-04-04
; PRIOR APPLICATION NUMBER: 07/062,925
; PRIOR FILING DATE: 1987-06-16
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 145
; TYPE: PRP
; ORGANISM: Bos taurus
US-09-940-601-3
```

```
Query Match          33.1%; Score 60; DB 9; Length 145;
Best Local Similarity 40.0%; Pred. No. 0.54;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
QY      3 MCTWDAFYLTVPQGVIEGVRHLVD 27
       :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDGVRKSD 48
```

```
RESULT 3
US-10-131-965-3
; Sequence 3, Application US/10131965
; Patent No. US20020165160A1
; GENERAL INFORMATION:
; APPLICANT: Whitehouse, Martha J.
; APPLICANT: Kavanaugh, Michael W.
; TITLE OF INVENTION: Angiogenically Effective Unit Dose of FGF and Method of
; FILE REFERENCE: 1296/121690505
; CURRENT APPLICATION NUMBER: US/10/131,965
; PRIOR FILING DATE: 2002-04-25
; PRIOR APPLICATION NUMBER: US/09/417,721
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/104,103
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 3
; LENGTH: 146
; TYPE: PRP
; ORGANISM: Human FGF-2
US-10-131-965-3
```

```
Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
QY      3 MCTWDAFYLTVPQGVIEGVRHLVD 27
       :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDGVRKSD 48
```

```
RESULT 4
US-10-131-965-5
; Sequence 5, Application US/10131965
; Patent No. US20020165160A1
; GENERAL INFORMATION:
; APPLICANT: Whitehouse, Martha J.
; APPLICANT: Kavanaugh, Michael W.
; TITLE OF INVENTION: Angiogenically Effective Unit Dose of FGF and Method of
; FILE REFERENCE: 1296/121690505
; CURRENT APPLICATION NUMBER: US/10/131,965
; CURRENT FILING DATE: 2002-04-25
```

```
; PRIOR APPLICATION NUMBER: US/09/417,721
; PRIOR FILING DATE: 1999-10-13
; PRIOR APPLICATION NUMBER: 60/104,103
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 146
; TYPE: PRP
; ORGANISM: bovine FGF-2
US-10-131-965-5
```

```
Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
QY      3 MCTWDAFYLTVPQGVIEGVRHLVD 27
       :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDGVRKSD 48
```

```
RESULT 5
US-10-108-195-2
; Sequence 2, Application US/10108195
; Publication No. US20030008820A1
; GENERAL INFORMATION:
; APPLICANT: Kwan, Chi-Pong
; APPLICANT: Venkataraman, Ganesh
; APPLICANT: Shriver, Zachary
; APPLICANT: Raman, Rahul
; APPLICANT: Sasisekharan, Ram
; TITLE OF INVENTION: Methods and Products Related to FGF Dimerization
; FILE REFERENCE: M00656/70076
; CURRENT APPLICATION NUMBER: US/10/108,195
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/279,165
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: Patentln version 3.1
; SEQ ID NO 2
; LENGTH: 146
; TYPE: PRP
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant of Native FGF2 with 9 N-terminal Residues Deleted
US-10-108-195-2
```

```
Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;
```

```
QY      3 MCTWDAFYLTVPQGVIEGVRHLVD 27
       :|:|:|:|:|:|:|:|:|:|:|:|
Db      24 YCKNGGFRLRHPDGRVDGVRKSD 48
```

```
RESULT 6
US-10-108-195-3
; Sequence 3, Application US/10108195
; Publication No. US20030008820A1
; GENERAL INFORMATION:
; APPLICANT: Kwan, Chi-Pong
; APPLICANT: Venkataraman, Ganesh
; APPLICANT: Shriver, Zachary
; APPLICANT: Raman, Rahul
; APPLICANT: Sasisekharan, Ram
; TITLE OF INVENTION: Methods and Products Related to FGF Dimerization
; FILE REFERENCE: M00656/70076
; CURRENT APPLICATION NUMBER: US/10/108,195
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/279,165
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 15
```



```

; SOFTWARE:PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant of Native FGF2 with 9 N-terminal Residues Deleted
US-10-108-195-3

Query Match      33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy      3 WCTMDAFYLTVPQGVIGVRHLVD 27
       :| | | | | | | | | | | | |
Db      24 YKNGCFFLRIHPDGRVDSGREKSD 48

RESULT 7
US-10-108-195-4
; Sequence 4, Application US/10108195
; Publication No. US20030008820A1
; GENERAL INFORMATION:
; APPLICANT: Kwan, Chi-Pong
; APPLICANT: Venkataraman, Ganesh
; APPLICANT: Shriver, Zachary
; APPLICANT: Raman, Rahul
; APPLICANT: Sasisekharan, Ram
; TITLE OF INVENTION: Methods and Products Related to FGF Dimerization
; FILE REFERENCE: M00656/70076
; CURRENT APPLICATION NUMBER: US/10/108,195
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/279,165
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 146
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Mutant of Native FGF2 with 9 N-terminal Residues Deleted
US-10-108-195-4

Query Match      33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches 10; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

Oy      3 WCTMDAFYLTVPQGVIGVRHLVD 27
       :| | | | | | | | | | | | |
Db      24 YKNGCFFLRIHPDGRVDSGREKSD 48

RESULT 8
US-10-108-195-7
; Sequence 7, Application US/10108195
; Publication No. US20030008820A1
; GENERAL INFORMATION:
; APPLICANT: Kwan, Chi-Pong
; APPLICANT: Venkataraman, Ganesh
; APPLICANT: Shriver, Zachary
; APPLICANT: Raman, Rahul
; APPLICANT: Sasisekharan, Ram
; TITLE OF INVENTION: Methods and Products Related to FGF Dimerization
; FILE REFERENCE: M00656/70076
; CURRENT APPLICATION NUMBER: US/10/108,195
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/279,165
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 146

```

```

      : TYPE: PRT
      : ORGANISM: Artificial Sequence
      : FEATURE:
      : OTHER INFORMATION: Native FGF2 with 9 N-terminal Residues Deleted
US-10-108-195-7

Query Match          33.1%; Score 60; DB 9; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches    10; Conservative   5; Mismatches    10; Indels     0; Gaps     0;

      OY       3 WCTWDAFYLTVHPGCVIEGRHLVD 27
            :|:|:|||||
      Db        24 YCKNGCFFLRHDPGRVGDGVREKSD 48

RESULT 9
US-09-802-365-2
: Sequence 2, Application US/09802365
: Patent No. US20020032153A1
: GENERAL INFORMATION:
: APPLICANT: Whitehouse, Martha Jo
: TITLE OF INVENTION: Methods and Compositions for the
: FILE REFERENCE: 1671.003
: CURRENT APPLICATION NUMBER: US/09/802,365
: PRIOR FILING DATE: 2001-03-09
: PRIOR APPLICATION NUMBER: 60/188,480
: PRIOR FILING DATE: 2000-03-10
: PRIOR APPLICATION NUMBER: 60/203,415
: PRIOR FILING DATE: 2000-05-11
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 146
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-802-365-4

Query Match          33.1%; Score 60; DB 10; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches    10; Conservative   5; Mismatches    10; Indels     0; Gaps     0;

      QY       3 WCTWDAFYLTVHPGCVIEGRHLVD 27
            :|:|:|||||
      Db        24 YCKNGCFFLRHDPGRVGDGVREKSD 48

RESULT 10
US-09-802-365-4
: Sequence 4, Application US/09802365
: Patent No. US20020032153A1
: GENERAL INFORMATION:
: APPLICANT: Whitehouse, Martha Jo
: TITLE OF INVENTION: Methods and Compositions for the
: FILE REFERENCE: 1671.003
: CURRENT APPLICATION NUMBER: US/09/802,365
: PRIOR FILING DATE: 2001-03-09
: PRIOR APPLICATION NUMBER: 60/188,480
: PRIOR FILING DATE: 2000-03-10
: PRIOR APPLICATION NUMBER: 60/203,415
: PRIOR FILING DATE: 2000-05-11
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 4
: LENGTH: 146
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-802-365-4

Query Match          33.1%; Score 60; DB 10; Length 146;
Best Local Similarity 40.0%; Pred. No. 0.55;
Matches    10; Conservative   5; Mismatches    10; Indels     0; Gaps     0;

```

```

QY      3 WCTWDAFYLLTVHPQGVIEGVRLVD 27
      :| | :| | | :| | | |
Db     24 YCKNGGFLLRIHPDGRVDGVREKSD 48

```

```

RESULT 11
US-09-771-302-2
; Sequence 2, Application US/09771302
; Patent No. US20020072489A1
; GENERAL INFORMATION:
; APPLICANT: Whitehouse, Martha J.
; TITLE OF INVENTION: Angiogenetically Effective Unit Dose of FGF-2 and Method
; TITLE OF INVENTION: of use
; FILE REFERENCE: 1296/12169US04
; CURRENT APPLICATION NUMBER: US/09/771,302
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: 09/385,114
; PRIOR FILING DATE: 1999-08-27
; PRIOR APPLICATION NUMBER: 60/104,102
; PRIOR FILING DATE: 1998-10-13
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 146
; TYPE: PR1
; ORGANISM: Bovis bovinus
; US-09-771-302-2

```

```
QY      3 WCITDAFYLLTVHPQGVIEGVRHLVD 27
      : | : | : | : | : | : |
Db     24 YCKNGGFELRIHPDGRVDGVREKSD 48
```

```

RESULT 12
US-09-886-856-2
; Sequence 2, Application US/09886856
; Patent No. US20020115603A1
; GENERAL INFORMATION:
; APPLICANT: Whitehouse, Martha Jo
; TITLE OF INVENTION: Methods and Compositions for the
; FILE REFERENCE: PP16090_004
; CURRENT APPLICATION NUMBER: US/09/886,856
; CURRENT FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: 60/213,504
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 60/264,572
; PRIOR FILING DATE: 2000-01-26
; PRIOR APPLICATION NUMBER: 60/276,549
; PRIOR FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 146
; TYPE: PR1
; ORGANISM: Bos taurus
; US-09-886-856-2

```

QY 3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
:|:|:|:|:|:|
Db 24 YCKNGGFRLRHPDGRVDGVREKSD 48

RESULT 13

```

US-09-886-856-4
: Sequence 4, Application US/09886856
: Patent No. US20020115603A1
: GENERAL INFORMATION:
: APPLICANT: Whitehouse, Martha Jo
: TITLE OF INVENTION: Methods and Compositions for the
: TITLE OF INVENTION: Treatment of Peripheral Artery Disease
: FILE REFERENCE: Ppl6090.004
: CURRENT APPLICATION NUMBER: US/09/886,856
: CURRENT FILING DATE: 2001-06-21
: PRIOR APPLICATION NUMBER: 60/213,504
: PRIOR FILING DATE: 2000-06-22
: PRIOR APPLICATION NUMBER: 60/264,572
: PRIOR FILING DATE: 2000-01-26
: PRIOR APPLICATION NUMBER: 60/276,549
: PRIOR FILING DATE: 2001-03-16
: NUMBER OF SEQ ID NOS: 9
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO. 4
: LENGTH: 146
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-886-856-4

```

```

QY      3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
      :| :| :| :| :| :|
Db     24 YCKNGGFFLRHPDGRVDGVREKSD 48

```

RESULT 14
 US-09-820-596-8
 Sequence 8, Application US/09820596
 Publication No. US20030022170A1
 GENERAL INFORMATION:
 APPLICANT: Khodadoust, Mehran Mohammad
 TITLE OF INVENTION: NOVEL FIBROBLAST GROWTH FACTORS AND
 THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
 NUMBER OF SEQUENCES: 18
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: FOLEY, HOAG & ELLIOT LLP
 STREET: One Post Office Square
 CITY: Boston
 STATE: MA
 COUNTRY: USA
 ZIP: 02109-2170
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/820,596
 FILING DATE: 29-Mar-2001
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: 09/036,594
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Arnold E., Beth
 REGISTRATION NUMBER: 35,430
 REFERENCE/DOCKET NUMBER: MIN-026.01
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 617-832-1000
 TELEFAX: 617-832-7000
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 147 amino acids
 TYPE: amino acid
 STRANDEDNESS: <unknown>
 TOPOLOGY: linear

```

;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-820-596-8

```

Query Match	33.18;	Score 60;	DB 9;	Length 147;
Best Local Similarity	40.08;	Pred. No. 0.55;		
Matches 10;	Conservative 5;	Mismatches 10;	Indels 0;	Gaps 0;

```

Oy      3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
          :| :| :| | :| | |
Db      25 YCKNGGFRLRHPDGRVDGVREKSD 49

```

```

RESULT 15
US-10-016-447-8
Sequence 8, Application US/10016447
Patent No. US20020090651A1
GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: Kirschner, Marc W.
TITLE OF INVENTION: Receptor-Ligand Assay
FILE REFERENCE: HU95-01A2
CURRENT APPLICATION NUMBER: US/10/016,447
CURRENT FILING DATE: 2001-12-10
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/08/776,207
PRIOR FILING DATE: EARLIER FILING DATE: 1997-06-23
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/441,629
PRIOR FILING DATE: EARLIER FILING DATE: 1995-05-15
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 08/279,217
PRIOR FILING DATE: EARLIER FILING DATE: 1994-07-22
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 150
TYPE: PRF
ORGANISM: Homo sapien
US-10-016-447-8

```

Query Match	33.1%;	Score 60;	DB 12;	Length 150;
Best Local Similarity	40.0%;	Pred. No. 0.56;		
Matches 10;	Conservative 5;	Mismatches 10;	Indels 0;	Gaps 0;

```

Qy      3 WCTWDAFYLTVHPQGVIEGVRHLVD 27
      .  :| :| :| :| :| :| :| :| :|
Db      33 YCKNGGFFLRHPDGRVDGVREKSD 57

```

Search completed: March 27, 2003, 10:20:35
Job time : 6.95122 secs

